



Facultad de Ciencias

Departamento de Biología Molecular

**Evolución del VIH-1 durante el proceso
de recuperación de la eficacia biológica
“in vitro”.**

Ramón Lorenzo Redondo

Madrid, 2011

UNIVERSIDAD AUTÓNOMA DE MADRID

Facultad de Ciencias

Departamento de Biología Molecular

**Evolución del VIH-1 durante el proceso de
recuperación de la eficacia biológica “in vitro”.**

Memoria presentada para optar al grado de Doctor en ciencias por:

Ramón Lorenzo Redondo

**Director de tesis:
Cecilio López Galíndez**

**Tutora:
M^a Ángeles Muñoz Fernández**

A mis padres y Ana con amor y gratitud.

“Durante la modificación de los descendientes y durante la incesante lucha de todas las especies por incrementar su número, cuanto más diversos lleguen a ser los descendientes, mayores serán sus posibilidades de triunfo en la lucha por la vida.”

Charles Darwin

El Origen de las Especies

“La unidad es la variedad, y la variedad en la unidad es la ley suprema del universo.”

Isaac Newton

AGRADECIMIENTOS

Una vez concluida esta fructífera, pero a la vez difícil etapa de mi vida, es hora de tener un pequeño detalle, mucho menor de lo que se merecerían, con todos aquéllos que me han apoyado y ayudado durante estos años, ya que como decía Quevedo “el agradecimiento es la parte principal de un hombre de bien”.

En primer lugar, quería agradecer a Cecilio su confianza en mí para realizar este trabajo y las oportunidades que me ha dado a lo largo de estos años. A pesar de todos los problemas que ha habido con el “*poderoso caballero*” y las dificultades pasadas, gracias a estas oportunidades, han resultado unos años enormemente productivos para mí. He disfrutado enormemente con nuestras charlas científicas y de otra muy diversa índole, acaecidas en esas innumerables tardes en el “labo”, así como de los viajes que hemos hecho juntos, en los que hemos compartido ciencia e incluso turismo, en algún pequeño resquicio que nos dejaba la primera. Finalmente, siempre le agradeceré el haberme introducido en el apasionante mundo de la virología y de la evolución, que creo marcarán mi futura carrera.

Es el turno para agradecer a mis compañeros o, mejor dicho, mi “otra” familia del Servicio de Virología Molecular todo lo que han hecho por mí en estos años. Haber compartido con ellos esta etapa ha sido fabuloso, ya que han hecho mucho más soportables los momentos difíciles y mucho mejores las épocas buenas. Normalmente, como frase hecha se suele decir que sin la ayuda de los que se nombran no hubiese sido posible terminar la tesis, pero en mi caso esta frase adquiere un sentido completo, ya que sin las concienzudas y, sobre todo, pacientes revisiones de María, Mónica, Concha e Isabel, no hubiese sido capaz de terminar esta tesis a tiempo. Además, he de agradecerles profundamente su trato conmigo, lo muchísimo que me han enseñado a lo largo de mi estancia en el laboratorio y la gran paciencia que han tenido ante mis incesantes dudas, estando siempre dispuestas a ayudarme. Probablemente sea muy difícil que en el futuro encuentre otro sitio, en el que se respire un ambiente familiar y de amistad como el que vosotras creáis en este laboratorio. También será difícil encontrar grandes amigos como los que han pasado a ser Carol, Cata y Víctor, con los que he compartido muy buenos momentos. He disfrutado muchísimo con la alegría y el buen humor que desprende Carol en todo momento, he aprendido mucho sobre la vida con el ejemplo que ha supuesto Cata para mí y he pasado instantes de gran diversión con mi “amiguete” Víctor. Asimismo, he de agradecerle a Rosa, que tanto nos cuida a todos, su inconmensurable ayuda. Tengo la certeza que de que la mayoría de las cosas no funcionarían en el “labo” si no fuese por ella.

No querría olvidarme tampoco, de los antiguos miembros del laboratorio como Carmen, que pasó los mismos momentos difíciles que yo, cuyo apoyo fue de gran ayuda, y Virginia, con quien hemos pasado tan buenos ratos. Por último, quisiera extender mi gratitud a Javi, Esther y Nuria del laboratorio del Dr. Alcamí con quienes hemos compartido esas agradables charlas, comidas y alguna que otra cena.

Dentro del Instituto de Salud Carlos III he de agradecer también a la Unidad de Genómica, ahora bajo la dirección de Ángel Zaballos, todo la labor que han hecho en el transcurso de esta tesis. Un simple

vistazo a la ingente cantidad de secuencias y análisis por “GeneScan” utilizados en este trabajo, dará una idea del enorme esfuerzo que han realizado Aurora, Ana, Miguel y Pilar. También agradecer la ayuda de Alicia en mis comienzos con el análisis de fragmentos.

En estos años, también me han ayudado en gran medida las diferentes colaboraciones que hemos realizado, en especial, ha sido un verdadero honor colaborar con Esteban Domingo, del que se puede aprender algo en cada una de las conversaciones que con él se mantienen, con Federico Morán, que nos ha introducido en el interesantísimo mundo de las redes neuronales artificiales, y Soledad Delgado, que tanto me ha ayudado con dichas redes y a la que pertenecen alguna de las figuras más bonitas de esta tesis. Además, me gustaría agradecer la invitación a lo largo de estos años a los seminarios llevados a cabo en el Centro de Astrobiología, que han supuesto una plataforma fabulosa de intercambio de ideas entre diversas áreas científicas relacionadas con nuestro campo.

Querría acabar la parte científica de estos agradecimientos con una mención a FIPSE, que ha financiado todo este trabajo y cuya labor ha sido, a mi entender, encomiable durante todos sus años de existencia.

Por otra parte, las primeras personas a las que tengo que agradecer cada segundo de mi vida son mis padres, Antonio y M^a Cruz. Ellos me han dado en cada momento todo lo que he necesitado, han sido mi mayor ejemplo y mis firmes valedores. Espero que alguna vez sea capaz de devolveros, al menos, una pequeña parte de todo lo que me habéis aportado. Una parte muy importante de todo este trabajo, es causa del gran esfuerzo que habéis realizado conmigo y, por eso, está dedicado a vosotros. También querría mencionar en estos agradecimientos a mi familia, en especial a Nieves, así como Belén, Lucía, Miguel y Javi, que tan bien me han acogido siempre, y tener un cariñoso recuerdo para mis abuelos, Íñigo y Marta, que espero hubiesen estado orgullosos de la tesis de su nieto. En mi familia también he de incluir, por supuesto, a mi gran amigo Antonio, cuya amistad ha sido un gran soporte siempre. Asimismo, tengo mucho que agradecer a Javi, un verdadero amigo siempre dispuesto a ayudar, a Moni y Francis, con quienes siempre he podido contar y que tanto me han ayudado todos estos años, a David, Elena, Juanqui y al resto de mis amigos, que han hecho estos años mucho mejores. Igualmente nombrar a la familia de Ana, agradeciendo su magnífico trato hacia mí, Sonia, Luis Miguel, Omi, Coto, Berta y, por supuesto, Rodrigo y a mi antiguo profesor José María, cuyas enseñanzas han forjado parte de la persona que ahora soy. Seguro que, debido a mi habitual despiste, me dejo a alguno en el tintero, espero sepan perdonarme.

Para terminar, mi último agradecimiento va dirigido a la que ha pasado a ser la persona más importante de mi vida, Ana. Tú has sido mi principal apoyo desde el momento en el que te conocí y siento que nunca hubiese sido capaz de llevar a cabo este trabajo si no te hubiese tenido a mi lado. Siempre has tenido una fe ciega en mí, a veces creo que demasiada, y, aunque tú no lo creas, has sido uno de mis principales modelos a seguir. Gracias por animarme siempre, ayudarme en todo y hacer mi vida mucho mejor. Nunca seré capaz de agradecerte todo lo que has hecho por mí.

LISTA DE ABREVIATURAS

³² P	Fósforo 32
ADN	Ácido desoxirribonucleico
AIC	Criterio de información de Akaike (Akaike information criterion)
aLRT	Método de razón de verosimilitud aproximado (approximate likelihood ratio test)
ANN	Redes neuronales artificiales (artificial neural networks)
ANOVA	Análisis de varianzas de una sola vía
ARN	Ácido ribonucleico
bmu	Unidad que mejor se aproxima (best matching unit)
dN	Sustituciones no sinónimas por sitios no sinónimos
dNTP	Desoxinucleótido
dS	Sustituciones sinónimas por sitios sinónimos
ECP	Efecto citopático
EDTA	Ácido etilendiaminotetraacético
G	Variación de tasas entre sitios
GTR	Modelo general time reversible
HBS	HEPES salino tamponado (HEPES buffered saline)
HKY	Modelo de sustitución de Hasegawa-Kishino-Yano
I	Proporción de sitios invariantes
LTR	Repetición terminal larga (long terminal repeat)
moi	Multiplicidad de infección (multiplicity of infection)
N	Sustituciones no sinónimas
N _e	Tamaño de población efectiva
NLS	Señal de localización nuclear (nuclear localization signal)
nt	Nucleótidos
ONUSIDA	Organización de la Naciones Unidas sobre el VIH/SIDA

ORF	Marco de lectura abierta (open Reading frame)
PBS	Sitio de unión al cebador (primer binding site)
PCR	Reacción en cadena de la polimerasa (polimerase chain reaction)
R^2	Cuadrado del coeficiente de regresión
R_n	Proporción del variante estudiado frente al virus de referencia
R_0	Proporción inicial del variante estudiado frente al virus de referencia
S	Sustituciones sinónimas
SBF	Suero bovino fetal
SDS	Dodecil sulfato sódico
SIDA	Síndrome de inmunodeficiencia adquirida
SOM	Mapas auto-organizativos (self organizing maps)
SSC	Citrato de sodio salino
$TCID_{50}$	Número de dosis infectivas 50
U	Unidades
UFP	Unidades formadoras de placas
VFA	Virus de la fiebre aftosa
VHC	Virus de la hepatitis humana C
VIH	Virus de la inmunodeficiencia humana
VIS	Virus de la inmunodeficiencia simia
VIS_{cpz}	Virus de la inmunodeficiencia simia de chimpancé
VIS_{sm}	Virus de la inmunodeficiencia simia de sooty mangabey
VSV	Virus de la estomatitis vesicular

Índice

RESUMEN.....	7
1. INTRODUCCIÓN	9
1.1 EL VIH.....	11
1.1.1 <i>La Pandemia del VIH-1: desde su descubrimiento a la situación actual.....</i>	<i>11</i>
1.1.2 <i>Clasificación del VIH-1.....</i>	<i>13</i>
1.1.3 <i>Estructura molecular.....</i>	<i>14</i>
1.1.4 <i>Ciclo biológico del VIH-1.....</i>	<i>18</i>
1.2 CONCEPTOS EVOLUTIVOS.....	20
1.2.1 <i>Modelo de cuasiespecies.....</i>	<i>20</i>
1.2.2 <i>Eficacia Biológica.....</i>	<i>22</i>
1.2.3 <i>Variabilidad.....</i>	<i>22</i>
1.2.4 <i>Tamaño poblacional.....</i>	<i>24</i>
1.2.5 <i>Selección natural.....</i>	<i>25</i>
1.2.6 <i>Paisaje de eficacias.....</i>	<i>26</i>
1.2.7 <i>Redes Neuronales Artificiales.....</i>	<i>27</i>
1.3 ANTECEDENTES.....	30
2. OBJETIVOS.....	31
3. MATERIALES Y MÉTODOS.....	35
3.1 CULTIVOS CELULARES.....	37
3.2 VIRUS E INFECCIONES.....	37
3.2.1 <i>Virus s61.....</i>	<i>37</i>
3.2.2 <i>Descripción de los clones virales D15, E11, G7, H12, I15, K15.....</i>	<i>37</i>
3.2.3 <i>Pases del virus en gran tamaño poblacional.....</i>	<i>38</i>
3.3 EFICACIA BIOLÓGICA DE LOS VIRUS.....	39
3.3.1 <i>Ensayo de eficacia biológica.....</i>	<i>39</i>
3.3.2 <i>Análisis por “GeneScan”.....</i>	<i>39</i>
3.4 CARACTERIZACIÓN GENÉTICA DE LOS VIRUS.....	40
3.4.1 <i>Obtención de las secuencias nucleotídicas.....</i>	<i>40</i>
3.4.1.1 <i>Extracción del ADN de las células de los cultivos infectados.....</i>	<i>40</i>
3.4.1.2 <i>Amplificación de las secuencias de los genomas virales.....</i>	<i>40</i>
3.4.1.3 <i>Secuenciación del genoma completo.....</i>	<i>41</i>
3.4.1.4 <i>Obtención de las cuasiespecies virales.....</i>	<i>42</i>
3.4.2 <i>Análisis de las secuencias nucleotídicas.....</i>	<i>44</i>
3.4.2.1 <i>Alineamiento de los genomas completos.....</i>	<i>44</i>
3.4.2.2 <i>Detección de las mutaciones y deleciones aparecidas.....</i>	<i>44</i>
3.4.2.3 <i>Análisis de la presión selectiva operando en los diferentes genes en el genoma completo.....</i>	<i>44</i>
3.4.2.4 <i>Estimación del modelo evolutivo seguido por las secuencias.....</i>	<i>45</i>
3.4.2.5 <i>Cálculo de la heterogeneidad de las cuasiespecies virales.....</i>	<i>45</i>
3.4.2.6 <i>Análisis de mutaciones sinónimas y no sinónimas dentro de las cuasiespecies.....</i>	<i>45</i>
3.5 ANÁLISIS FILOGENÉTICOS DE LAS SECUENCIAS VIRALES OBTENIDAS.....	46
3.5.1 <i>Estimación filogenética por Máxima Verosimilitud.....</i>	<i>46</i>

3.5.2 Análisis de las tasas de evolución de los distintos virus mediante estimación bayesiana.....	46
3.5.3 Redes filogenéticas de las cuasiespecies virales.....	47
3.6 CARACTERIZACIÓN FENOTÍPICA DE LOS VIRUS OBTENIDOS EN LOS CULTIVOS CELULARES.	47
3.6.1 Titulación de los virus.....	47
3.6.2 Determinación de los niveles de proteína p24 viral.....	47
3.6.3 Ensayo de la citopatogenicidad de los virus estudiados.....	47
3.6.4 Actividad Retrotranscriptasa.....	48
3.7 ANÁLISIS POR REGRESIÓN MÚLTIPLE.....	48
3.8 ANÁLISIS DEL PAPEL DE CAMBIOS DE AMINOÁCIDO APARECIDOS EN EL PROCESO DE PÉRDIDA Y RECUPERACIÓN DE EFICACIA BIOLÓGICA MEDIANTE MUTAGÉNESIS DIRIGIDA.....	49
3.8.1 Obtención de secuencias de ADN con mutaciones aparecidas en el proceso de pérdida y recuperación de eficacia biológica.....	49
3.8.2 Generación y obtención de los virus mutados.....	51
3.8.3 Modelización de las estructuras mutadas.....	51
3.9 GENERACIÓN DE PAISAJES ADAPTATIVOS.....	52
3.9.1 Obtención de un paisaje adaptativo por técnicas filogenéticas.....	52
3.9.2 Obtención de un paisaje adaptativo utilizando redes neuronales.....	52
3.10 ANÁLISIS ESTADÍSTICOS.....	54
4. RESULTADOS.....	57
4.1 ANÁLISIS DE LA RECUPERACIÓN DE LA EFICACIA BIOLÓGICA DE CLONES BIOLÓGICOS DEL VIH-1.....	59
4.1.1 Crecimiento continuo de la eficacia biológica global en clones biológicos del VIH-1 durante 30 pases seriados en gran tamaño poblacional.....	59
4.2 ANÁLISIS GENÉTICOS DE LOS VIRUS ESTUDIADOS.....	62
4.2.1 Análisis de los genomas completos globales de los virus.....	62
4.2.1.1 Estudio de las mutaciones aparecidas a lo largo del proceso de recuperación de eficacia biológica.....	62
4.2.1.2 Aumento de las posiciones heterogéneas a lo largo de los pases de recuperación.....	68
4.2.1.3 Aparición de deleciones en la región V1-V2 del gen env.....	69
4.2.2 Análisis de las cuasiespecies virales.....	70
4.2.2.1 Aumento de la heterogeneidad de las cuasiespecies a lo largo de los pases de recuperación.....	70
4.2.2.2 Existencia de una correlación global significativa entre heterogeneidad y eficacia biológica a lo largo de los pases de recuperación.....	71
4.2.2.3 Análisis de las fuerzas selectivas operando en las cuasiespecies durante la recuperación de eficacia biológica.....	73
4.3 ESTUDIOS FILOGENÉTICOS.....	75
4.3.1 Modelo evolutivo seguido por los virus en los pases en gran tamaño poblacional.....	75
4.3.2 Estimación de la filogenia de los virus en el proceso de recuperación por Máxima Verosimilitud.....	75

4.3.3 Cálculo de la distancia de cada virus a la raíz del árbol estimado por Máxima Verosimilitud.....	77
4.3.4 Cálculo de las tasas de evolución de los virus estudiados y su relación con la recuperación de eficacia.	78
4.3.5 Análisis por redes de la evolución de las cuasiespecies virales.	80
4.4 CARACTERIZACIÓN FENOTÍPICA.	84
4.4.1 Estudio de la citopatogenicidad de los virus a lo largo de la recuperación de la eficacia biológica.	84
4.4.2 Determinación del título viral y de la producción de proteína p24.....	86
4.4.3 Análisis la actividad RT a lo largo de los pases de recuperación.	87
4.5 ANÁLISIS MULTIVARIABLE POR REGRESIÓN LINEAL DE LA RECUPERACIÓN DE LA EFICACIA BIOLÓGICA.	89
4.6 ESTUDIO DE MUTACIONES APARECIDAS EN LOS EXPERIMENTOS DE EVOLUCIÓN “IN VITRO” CON DIFERENTES TAMAÑOS POBLACIONALES.	92
4.6.1 Análisis de cambios de aminoácidos aparecidos durante los experimentos de evolución “in vitro”, modelado por homología de sus estructuras y caracterización fenotípica de los virus generados.	92
4.6.1.1 Mutaciones aparecidas en los virus durante los pases placa a placa.....	93
4.6.1.2 Mutaciones aparecidas en los virus al realizar la semilla posterior a los pases placa a placa.....	97
4.6.1.3 Mutaciones aparecidas en los virus a lo largo de los pases de recuperación de eficacia biológica.	98
4.6.2 Resumen de los resultados obtenidos en el estudio de mutantes.	101
4.7 GENERACIÓN DE UN PAISAJE DE EFICACIAS A PARTIR DE DATOS REALES Y SU APLICACIÓN PARA EL ESTUDIO DEL PROCESO DE RECUPERACIÓN DE EFICACIA DEL VIH-1 Y LA ESTIMACIÓN DE EFICACIAS DESCONOCIDAS.	102
4.7.1 Generación de un paisaje de eficacias reales utilizando métodos filogenéticos.	102
4.7.2 Generación de un paisaje de eficacias real predictivo utilizando redes neuronales artificiales.	105
5. DISCUSIÓN	109
5.1 ANÁLISIS DE LA RECUPERACIÓN DE EFICACIA BIOLÓGICA.....	111
5.2 ANÁLISIS GENÉTICO DEL PROCESO DE RECUPERACIÓN DE EFICACIA BIOLÓGICA.	113
5.3 FUERZAS SELECTIVAS OPERANDO DURANTE LOS PASES DE RECUPERACIÓN.....	115
5.4 FENOTIPO DE LOS VIRUS DURANTE LA RECUPERACIÓN DE LA EFICACIA BIOLÓGICA. .	116
5.5 MUTACIONES APARECIDAS DURANTE LA PÉRDIDA Y RECUPERACIÓN DE EFICACIA “IN VITRO” DEL VIH-1 ESTUDIADAS POR MUTAGÉNESIS DIRIGIDA.....	117
5.6 GENERACIÓN DEL PAISAJE DE EFICACIAS EXPERIMENTAL DEL PROCESO DE RECUPERACIÓN DE EFICACIA “IN VITRO” DEL VIH-1.	119
5.7 DISCUSIÓN GENERAL.	120
6. CONCLUSIONES.....	121
7. BIBLIOGRAFÍA.....	125
TRABAJOS PUBLICADOS.....	137

RESUMEN

Human immunodeficiency virus type 1 (HIV-1) constitutes one of the major medical challenges due to infectious diseases the world is facing, affecting more than 30 million people. Due to its high mutation rate, enormous population sizes and rapid turnover, HIV-1, as the other RNA viruses, generates a swarm of mutants known as viral quasispecies. The quasispecies and not single genomes are the object of evolution, thus studying its properties will lead to a better understanding of viral evolution. For that purpose, “*in vitro*” experiments are usually employed to simulate the population size changes the virus suffers during natural infections.

In the present thesis, we have subjected 6 biological clones, that had suffered drastic fitness losses due to plaque-to-plaque passages, to 30 large population passages in cell culture to recover fitness. After the passages, we performed a complete analysis of the viruses including fitness, phenotypic properties, nucleotide sequences at the consensus level and quasispecies heterogeneity. We also studied individually some of the mutations that appeared during the passages performed. Finally, we used all the nucleotide sequences and obtained along the passages and their fitness values to generate a real fitness landscape of the recovery process of the viruses.

All viral populations recovered fitness after the 30 passages with a continuing overall grow (from a mean fitness of 0.52 in passage 1 to 1.58 in passage 31), but with significant differences between viruses. When the potential causes of fitness recovery were analysed, a high positive correlation between heterogeneity and fitness ($p\text{-value} = 0.0057$) was obtained. The quasispecies heterogeneity together with the evolutionary rate were the variables most associated with the fitness recovery using a multiple regression analysis.

We also observed that this increase in quasispecies complexity allowed the virus to fix new beneficial mutations, of which almost 90% were non-synonymous, with an average of 8.3 mutations per virus at passage 31. The mutational pattern, although presenting some convergent mutations and deletions, was very different between viruses and did not present hotspots. Interestingly, some mutations associated with important phenotypic effects, such as tropism change or antiviral resistance, appeared in these “*in vitro*” conditions without any external selective pressure.

In addition, we detected that not only the consensus sequences were positively selected but also, at the quasispecies level, the mutant spectra was shaped by positive selection. Using all quasispecies sequences within each virus, we generated phylogenetic networks to study the appearance of mutations, to detect the imposition of beneficial changes in viral populations and to describe the quasispecies dynamics throughout the passages.

Ultimately, we generated a real fitness landscape with all the sequences obtained during these analyses, using phylogenetic methods that produced descriptive landscapes. Moreover, we used artificial neural networks (ANN) to gain projective capacity, allowing the construction of predictive fitness landscapes that could be used to estimate unknown fitness values.

Using all the analyses performed in this thesis, we have disclosed different pathways of “*in vitro*” fitness recovery at the consensus level and analysed the dynamics of variability generation and mutant selection within viral quasispecies. To conclude, in the process of HIV-1 “*in vitro*” biological fitness recovery the quasispecies heterogeneity is one of the most important factors driving viral evolution. The generation of complexity leads to the fixation of beneficial mutations and determines the properties of the viral populations.

1. Introducción

1.1 El VIH.

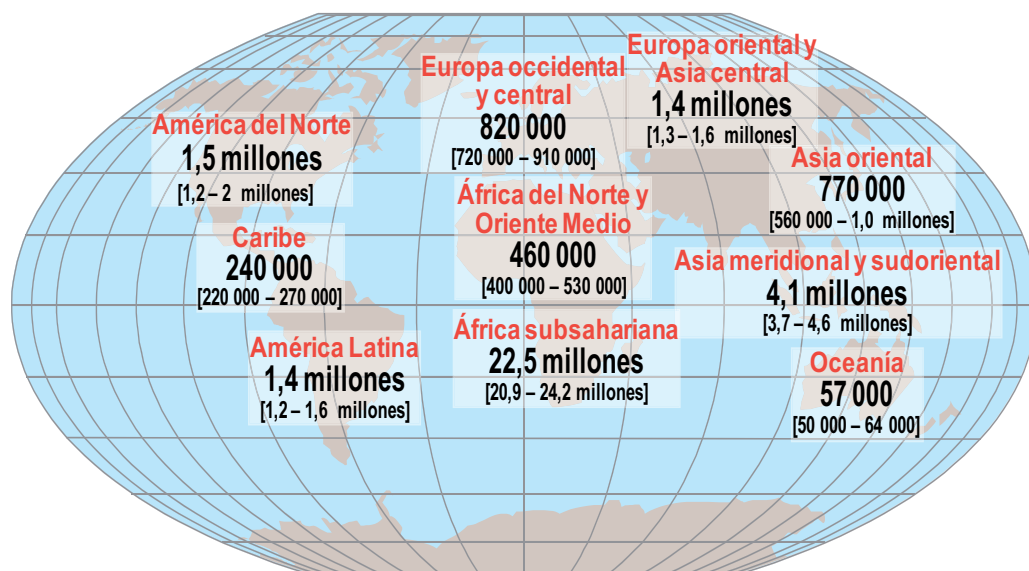
1.1.1 La Pandemia del VIH-1: desde su descubrimiento a la situación actual.

El síndrome de inmunodeficiencia adquirida (SIDA) es una enfermedad que afecta al sistema inmune causada por el virus de la inmunodeficiencia humana (VIH) y que actualmente constituye una epidemia de dimensiones mundiales. El SIDA fue reconocido como una nueva entidad clínica en 1981 (Gottlieb y cols., 1981; Masur y cols., 1981). Este síndrome se caracteriza por la presencia de anomalías inmunológicas generalmente acompañadas por infecciones oportunistas, desórdenes neurológicos y formas inusuales de cáncer. Los primeros estudios epidemiológicos demostraron que este nuevo agente infeccioso se transmitía a través de la sangre o sus derivados, a través de los fluidos genitales y verticalmente de madre a hijo. Dos años más tarde, en 1983, se identificó el agente causal de la enfermedad, al aislarse por primera vez a partir de nódulos linfáticos de un paciente con linfadenopatía el virus de la inmunodeficiencia adquirida tipo 1 (VIH-1) (Barre-Sinoussi y cols., 1983). Un año después, este virus se detectó en muestras de suero de pacientes con SIDA (Gallo y cols., 1984).

La secuencia más antigua de VIH-1 procede de una muestra de plasma del Zaire del año 1959 (Zhu y cols., 1998). El origen del virus ha suscitado bastante controversia, siendo muy amplio el rango de datación entre los diferentes trabajos. Las dataciones más aceptadas sitúan el origen del VIH a finales del siglo XIX o principios del XX (Korber y cols., 2000; Worobey y cols., 2008), aunque algún estudio ha llegado a situar dicho origen en el siglo XVII (Salemi y cols., 2001). Mayor consenso existe con respecto al origen de la pandemia actual, que estudios recientes sitúan en torno a 1930 (Worobey y cols., 2008), mientras que el origen de la epidemia en Norteamérica y Europa se ha datado en torno a 1969 (Gilbert y cols., 2007).

Según el informe del año 2010 del Programa Conjunto de las Naciones Unidas sobre el VIH/SIDA (ONUSIDA) (http://www.unaids.org/globalreport/Global_report_es.htm), el crecimiento de la epidemia global de SIDA parece haberse estabilizado, pero sigue constituyendo un enorme problema económico y social. El número anual de nuevas infecciones ha ido cayendo constantemente desde finales de los años 90 y las muertes relacionadas con el SIDA son menores debido a los avances en el tratamiento antirretroviral. Aunque el número de nuevas infecciones ha caído, este número globalmente sigue siendo alto y, debido a la disminución de la mortalidad, la prevalencia de la infección ha aumentado mundialmente. En 2009, la epidemia de SIDA causó 1,8 millones de muertes y se estima que se produjeron unas 2,6 millones de nuevas infecciones, elevándose la cifra total de personas viviendo con VIH a 33,3 millones alrededor del globo. Desde el comienzo de la epidemia, el África subsahariana es, con mucho, la región más afectada del mundo. En ésta región vive el 10% de la población mundial pero acumula el 68% del total de personas infectadas con el VIH (22,5 millones, con un 5% de prevalencia en la población adulta), siendo además la que acumula el mayor número de nuevos casos (1,8 millones, el 69% del total). En la Figura 1.1 se puede observar la distribución de la epidemia en el mundo.

Número estimado de adultos y niños infectados con el VIH en 2009



Total: 33,3 millones (31,4 – 35,3 millones)

Figura 1.1 Número de adultos y niños infectados con el VIH en 2009. Figura de ONUSIDA del año 2009 adaptada con los datos estimados en el informe anual de 2010. Entre paréntesis se muestra el rango de error de la estimación.

Los datos arrojan una disminución tanto en el número de nuevas infecciones como en la mortalidad causada por el VIH. Esta disminución ha sido principalmente debida al considerable incremento en el uso del tratamiento antirretroviral en todo el mundo. Aunque el tratamiento sigue siendo mayoritario en los países desarrollados, en el año 2009 se produjo un aumento en el número de personas en tratamiento desde los 1,2 millones, hasta alcanzar los 5,5 millones de personas tratadas, constituyendo un 30% de incremento con respecto a 2008. De este incremento la mayoría se produjo en Europa del Este, el África subsahariana, Asia y el Caribe, lo que indica que los esfuerzos por extender el tratamiento en los países menos desarrollados están dando sus frutos. Sin embargo, a pesar de estas mejoras, tan sólo el 36% de la población que lo necesita recibe tratamiento y la enfermedad no está aún controlada en algunos países o regiones.

La infección por VIH-1 sigue siendo una de las primeras causas de mortalidad mundial y su incidencia, especialmente en países del África subsahariana, sigue teniendo niveles muy elevados. Estos datos demuestran que es necesario seguir aumentando el uso y variedad de antirretrovirales para controlar la pandemia, así como políticas de salud pública relacionadas con la prevención, que han demostrado ser muy eficaces. Por otro lado, la erradicación del virus en los individuos infectados sólo parece posible con el desarrollo de una vacuna efectiva, por lo que el estudio y comprensión de las características y evolución del virus son indispensables para alcanzar este objetivo.

1.1.2 Clasificación del VIH-1.

Sobre la base de sus características genéticas, morfológicas y biológicas, el VIH-1 se clasifica dentro del género *Lentivirus* (Ratner, Gallo y Wong-Staal, 1985; Wain-Hobson y cols., 1985) perteneciente a la familia *Retroviridae*. Se han descrito dos tipos de virus que causan el SIDA, el VIH-1 y el VIH-2, cuyo origen se supone que fue debido a eventos independientes de salto de especie del virus de la inmunodeficiencia simia (VIS) al hombre. Se postula que el tipo 1 proviene del VIS de chimpancé (VIS_{cpz}), mientras que el VIH-2 proviene del VIS de mangabeye fuliginoso (sooty mangabey monkey), VIS_{sm} (Heeney, Dalglish y Weiss, 2006).

El VIH-1 se divide en cuatro grupos: M (responsable de la pandemia mundial), O (endémico de Camerún y de otros países del África central occidental), N (que fue descrito en 7 pacientes de Camerún (Simon y cols., 1998)) y P (recientemente descubierto también en Camerún (Plantier y cols., 2009)). Actualmente se cree que la existencia de estos cuatro grupos es debida a distintas introducciones en humanos del VIS de primates (Nerrienet y cols., 2005).

Dentro del grupo M se distinguen nueve subtipos A, B, C, D, F, G, H, J y K, con una distribución geográfica característica, aunque todos se encuentran representados en el África subsahariana. Estos subtipos se definen en base a la distancia genética entre las secuencias del gen de la envuelta, considerándose distintos subtipos cuando la divergencia entre ellos es mayor del 25%. Por su parte, la diferencia entre grupos distintos es mayor del 35%. Además se han descrito numerosas formas recombinantes (en la actualidad 49 descritas, <http://www.hiv.lanl.gov/content/sequence/HIV/CRFs/CRFs.html>).

La enorme diversidad generada por el VIH no sólo se hace patente al comparar entre subtipos y grupos, sino también en la distancia genética observada dentro de un mismo virus en un paciente, que se ha calculado entre el 2% y el 10% (Myers, 1994). Al comparar estos valores con otros virus ARN, por ejemplo el virus de la gripe, se puede tener una idea de la magnitud de esta variabilidad. Así, la diversidad generada dentro de un único paciente puede llegar a ser mayor que la generada en una epidemia global del virus de la gripe (Figura 1.2) (Korber y cols., 2001; Walker y Burton, 2008). Esta enorme variabilidad que presenta el VIH constituye el factor fundamental responsable de su capacidad de evasión del sistema inmune, además de haber sido asociada con la falta de éxito de las vacunas y, por lo tanto, con su imposible (hasta el momento) erradicación. Así, el estudio de la generación de dicha variabilidad y de sus propiedades resulta fundamental para la progresión de los programas de vacunación y el control de la infección.

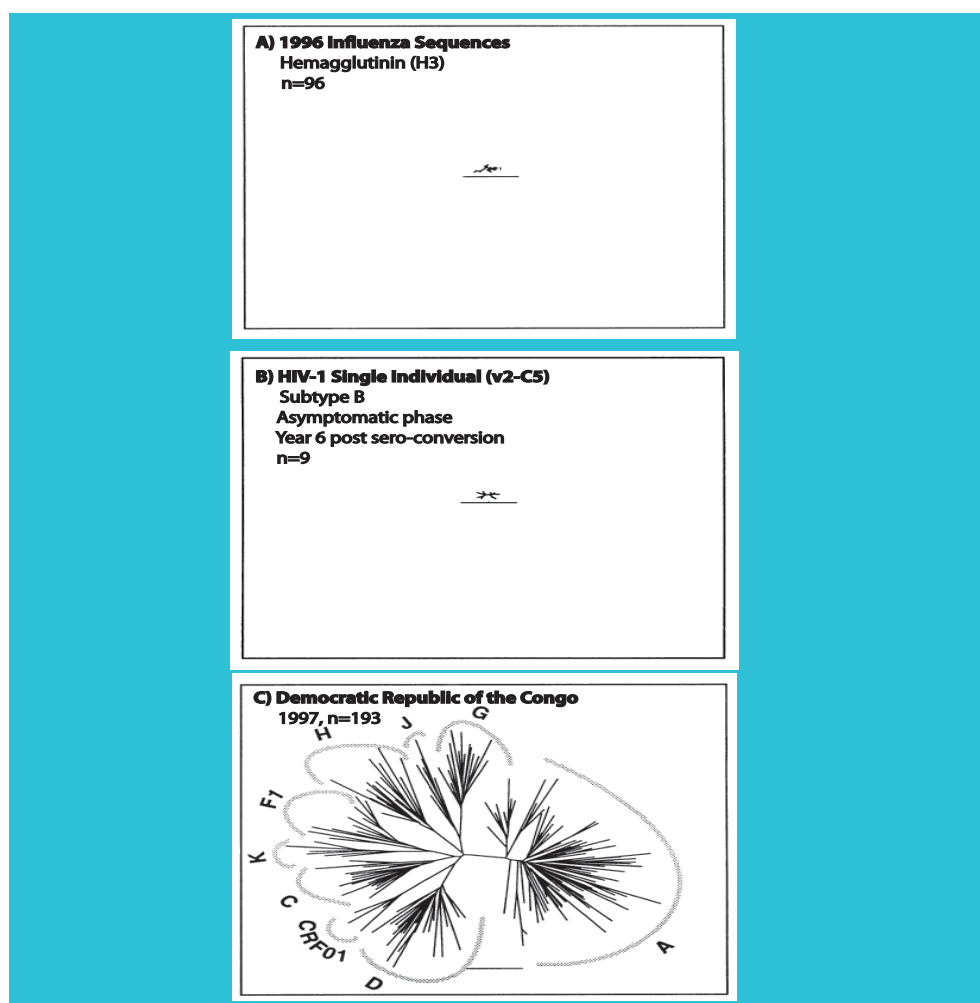


Figura 1.2 Comparación de las distancias evolutivas de secuencias en la región V2-C5 de la envuelta de VIH-1 y el dominio H3 del gen de la Hemagglutina del virus de la gripe. Se muestran los árboles correspondientes a: A) Todas las secuencias del año 96 de la base de datos de gripe de Los Alamos. B) Secuencias obtenidas de un único paciente asintomático de VIH-1, recogidas en el mismo momento después de 6 años de infección. C) Todas las muestras recogidas en el año 97 en la República Democrática del Congo. Figura obtenida de Korber y cols 2001. En todos los árboles la n constituye el número de pacientes incluidos en el análisis.

1.1.3 Estructura molecular.

El examen por microscopía electrónica revela que el virión del VIH-1 tiene una estructura esférica de 80 a 130 nm de diámetro constituida por tres capas concéntricas. La capa interna (nucleoide o core), constituida por la proteína de la cápside (CA o p24), tiene forma de cono truncado y contiene dos copias de ácido ribonucleico (ARN) viral de cadena sencilla y polaridad positiva (Hu y Temin, 1990) de aproximadamente 9,2 Kb, formando un complejo ribonucleoproteico constituido fundamentalmente por las proteínas retrotranscriptasa (RT o p66/51), integrasa (IN o p32) y la proteína de la nucleocápside (NC o p7), así como por la proteína Vpr. La capa intermedia forma una cápside icosaédrica, que está constituida por la proteína de la matriz (MA o p17). La capa externa o envuelta está constituida por una bicapa lipídica derivada de la célula infectada que contiene proteínas celulares y por espículas formadas por oligómeros de las glicoproteínas de la envuelta vírica: la glicoproteína externa gp120 y la glicoproteína transmembrana gp41. Ambas proteínas se encuentran asociadas no covalentemente en la superficie del virión. En la Figura 1.3 se muestra una representación de la estructura del virión del VIH-1.

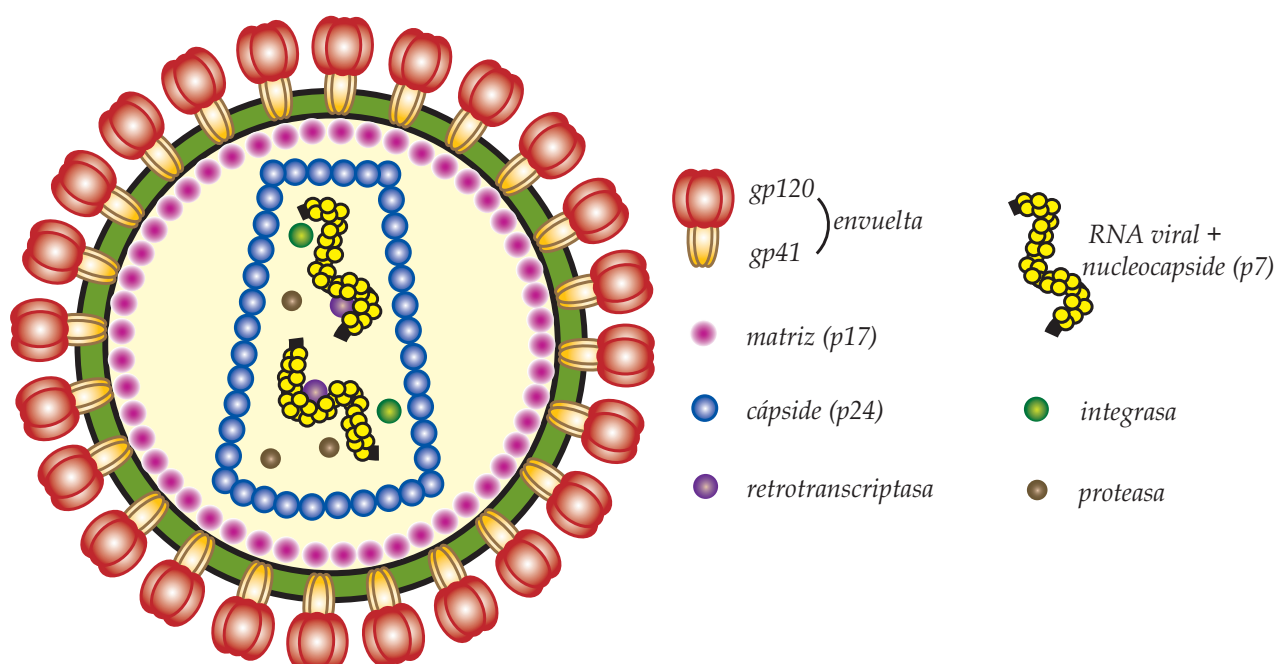


Figura 1.3 Representación esquemática de un virión maduro de VIH-1. Figura cedida por la Dra. C. Reis.

El genoma del VIH-1 tiene 9 marcos de lectura abierta (open reading frames, ORF). Estos ORFs están flanqueados por repeticiones terminales de varios cientos de nucleótidos denominadas repeticiones terminales largas (long terminal repeat, LTR). Tres de los ORF corresponden a los genes estructurales virales, *gag*, *pol* y *env*, y los seis ORF restantes corresponden a los genes reguladores, *tat* y *rev*, y a los genes accesorios, *vif*, *nef*, *vpr* y *vpu*.

Cada uno de estos ORF da lugar directamente a proteínas virales por procesamiento, a excepción de los 3 genes estructurales, que codifican tres precursores poliproteicos que son posteriormente procesados por proteasas celulares o virales para formar las proteínas estructurales maduras. Estas tres poliproteínas son Gag (Pr55), Gag-Pol (Pr160) y Env (Pr160). Gag da lugar a la proteína p17, la proteína p24, la proteína p7, la proteína p6 y las proteínas p1 y p2. Un deslizamiento ribosomal del marco de lectura entre *gag* y *pol*, da lugar al precursor de Gag-Pol, cuyo procesamiento da lugar a la proteasa (PR, p10), la RT, formada por un heterodímero de las proteínas p66 y p51, y la IN. Finalmente, el precursor Env da lugar a la gp120 y la gp41.

Cada proteína cumple su función en la infección viral, incluso, debido a la optimización a la que la evolución ha llevado a este virus, muchas de estas proteínas realizan varias funciones en diferentes procesos de la infección. A continuación se enumeran brevemente las funciones descritas hasta la fecha de la diferentes proteínas virales:

Gen	Proteína	Función
Genes Estructurales		
<i>gag</i>	p17	Constituye la matriz, que se sitúa en el interior y alrededor de la cápside viral. Esta proteína media en el proceso de ensamblaje del virión (Freed, Englund y Martin, 1995; Yuan y cols., 1993; Zhou y cols., 1994), en el proceso de gemación del virus (Göttlinger, 2001) y, además, dirige al complejo de pre-integración hacia el núcleo celular (Bukrinsky y cols., 1993; von Schwedler, Kornbluth y Trono, 1994).
<i>gag</i>	p24	Forma la cápside viral (Göttlinger, 2001).
<i>gag</i>	p7	Está acomplejada con el ARN viral formando el denominado complejo ribonucleoproteico (Berkowitz, Fisher y Goff, 1996).
<i>gag</i>	p6	Favorece el ensamblaje del virión y su liberación de la membrana plasmática (Gottlinger y cols., 1991; Huang y cols., 1995; Zhu, Jian y Zhao, 2004).
<i>pol</i>	PR	Proteasa viral que procesa los precursores poliproteicos de las proteínas Gag y Pol en sus componentes proteicos simples (Crawford y Goff, 1985; Katoh y cols., 1985).
<i>pol</i>	RT	Polimerasa viral que convierte el ARN de cadena sencilla en ADN de doble cadena (Baltimore, 1970; Temin y Mizutani, 1970).
<i>pol</i>	IN	Enzima que media la integración del ADN viral en el genoma de la célula hospedadora (Brown y cols., 1989; Fujiwara y Mizuuchi, 1988).
<i>env</i>	gp120	Glicoproteína de superficie, implicada en la unión y fusión con la membrana celular de la célula hospedador (White, 1992). Interacciona con el receptor celular CD4 (Sattentau y cols., 1993; Thali y cols., 1993).
<i>env</i>	gp41	Proteína transmembrana, implicada en la unión y fusión con la membrana celular de la célula hospedador (White, 1992).

Genes Reguladores		
<i>tat</i>	Tat	Incrementa enormemente los niveles de transcripción viral por interacción con un motivo, denominado TAR, en el ARN viral en células infectadas (Dayton y cols., 1986; Fisher y cols., 1986).
<i>rev</i>	Rev	Es la proteína viral que media el transporte de los mensajeros de ARN virales al citoplasma (Meyer y Malim, 1994).
Genes Accesorios		
<i>vif</i>	Vif	Proteína también denominada factor viral de infectividad. Contrarresta el efecto de la citidin deaminasa APOBEC3G, que es un inhibidor de la infección viral (Kremer y Schnierle, 2005; Malim, 2009; Schrofelbauer, Yu y Landau, 2004), siendo necesaria para la generación de viriones infectivos en células no permisivas (Strebel y cols., 1987).
<i>vpr</i>	Vpr	Estimula la transcripción de los genes virales, facilita el transporte al núcleo del complejo de preintegración (Heinzinger y cols., 1994; Lu, Spearman y Ratner, 1993) y mantiene a la célula infectada en la fase G2 del ciclo celular (Poon y cols., 1998).
<i>vpu</i>	Vpu	Facilita la liberación de virus por la interacción con la teterina celular CD317 (Neil, Zang y Bieniasz, 2008) y la degradación del receptor CD4 (Bour, Schubert y Strebel, 1995; Chen y cols., 1993).
<i>nef</i>	Nef	Reduce la expresión en la superficie de la célula infectada del CD4 (Garcia y Miller, 1991) y del complejo mayor de histocompatibilidad I (MHC I), (Schwartz y cols., 1996). También incrementa la infectividad viral (Chowers y cols., 1994; Miller y cols., 1994; Spina y cols., 1994), y la eficiencia de la retrotranscriptasa viral (Schwartz y cols., 1995).

1.1.4 Ciclo biológico del VIH-1.

El primer paso del ciclo de infección del VIH-1 es la unión de la proteína gp120 con el receptor CD4 localizado en la superficie de los linfocitos T colaboradores y en células de la estirpe monocito-macrófago. Además, para la entrada a la célula, el virus necesita unirse también a un co-receptor, constituido por receptores de quimiocinas (Paso 1 de Figura 1.4). Estos co-receptores son, fundamentalmente, el receptor de fusina CXCR4, cuyo ligando natural es SDF-1, y el receptor CCR5 de las quimioquinas Rantes, MIP-1 α y MIP-1 β (Choe y cols., 1996; Deng y cols., 1996; Doranz y cols., 1996; Feng y cols., 1996). La unión a uno u otro co-receptor define el tropismo del virus. Una vez realizada esta unión entre la proteína gp120 y sus receptores (Paso 1), se produce la fusión entre la membrana viral y la celular, que permite la internalización de la nucleocápside (Paso 2). En este momento, la cápside se abre parcialmente y se produce la retrotranscripción del genoma viral por parte de la RT viral (Paso 3). Dicha RT convierte el ARN vírico en ácido desoxirribonucleico (ADN) de doble hebra, con una LTR en cada extremo; este ADN es incorporado en un complejo de pre-integración y transportado activamente al núcleo celular (Paso 4). En el núcleo, la integrasa viral media la integración del ADN viral en el genoma de la célula hospedadora, constituyéndose el provirus integrado. Esta es una estructura estable que permanece indefinidamente en el genoma celular.

Cuando está integrado, el virus puede permanecer latente, replicarse de forma limitada o experimentar una replicación masiva, según el estado de activación de la célula, con el consiguiente efecto citopático sobre la célula infectada. La transcripción de los genes virales se lleva a cabo por la ARN polimerasa II de la célula cuando ésta recibe señales de activación, generando un único precursor ARN que es utilizado como genoma o como ARN mensajero. En este paso, la proteína viral Tat actúa favoreciendo la alta producción de los transcritos virales, que son transportados al citoplasma gracias a la proteína viral Rev (Paso 5). El precursor de la proteína de la envuelta es traducido en el retículo endoplasmático y procesado proteolíticamente en el aparato de Golgi, mientras que las poliproteínas gag y gag-pol son traducidas en los polirribosomas (Paso 6). Tanto las proteínas de la envuelta como las de los genes virales *gag* y *pol*, son transportadas a la membrana celular, donde, junto al ARN viral generado, comienzan a formar complejos virales en la membrana citoplasmática, dando lugar a partículas esféricas virales inmaduras que contienen las glicoproteínas de la envuelta gp120 y gp41 (Paso 7). Finalmente, estas partículas se liberan de la membrana plasmática (Paso 8). La liberación de la partícula viral de la célula hospedadora produce la maduración de la partícula viral, mediante el procesamiento proteolítico de las poliproteínas Gag y Pol por la proteasa viral, dando lugar a la forma característica en cono truncado de la cápside madura, que constituye el virión infeccioso (Paso 9).

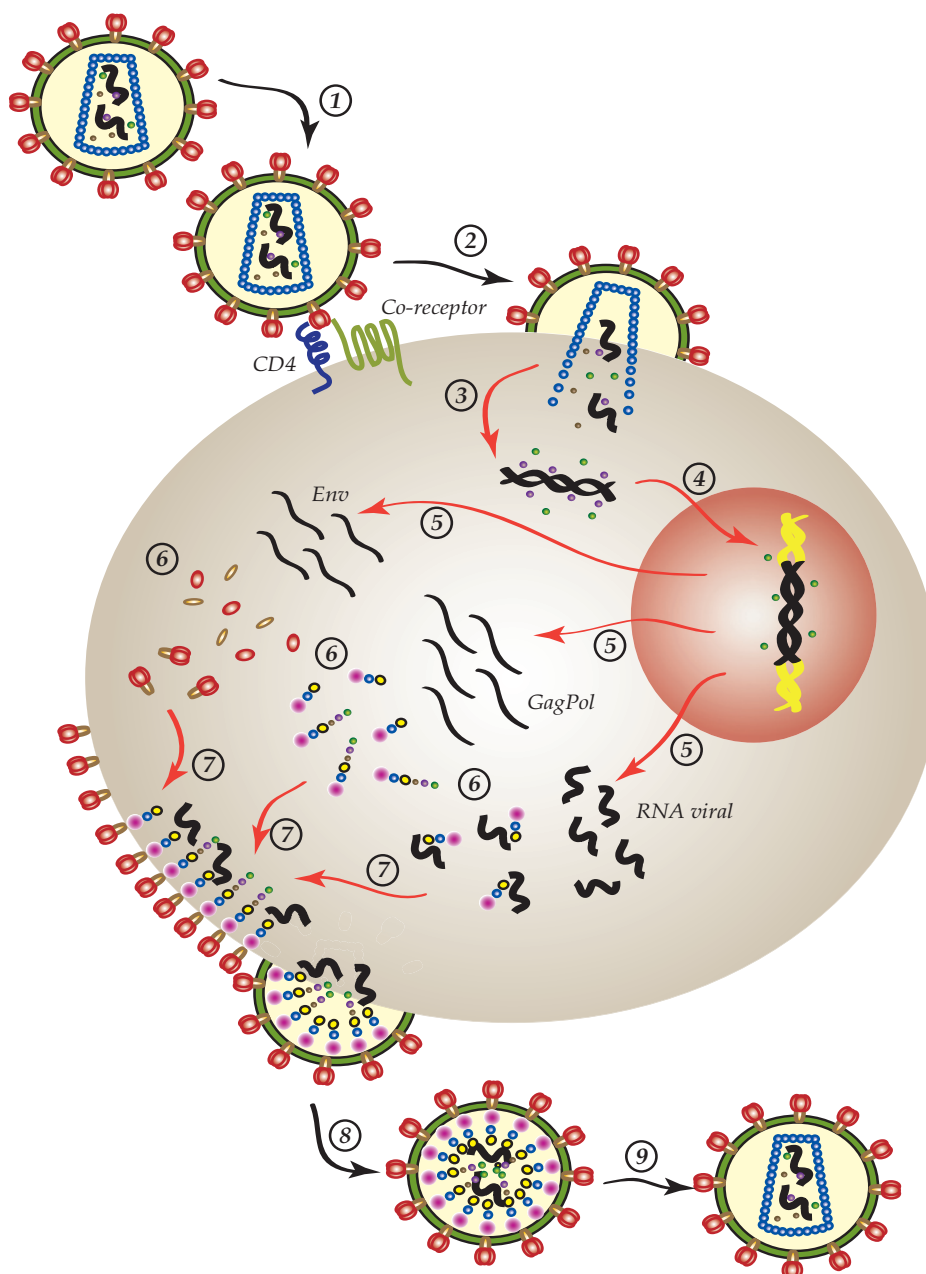


Figura 1.4 Ciclo biológico del VIH-1. Representación esquemática del ciclo biológico del VIH-1. (1) Unión del virus a los receptores celulares. (2) Fusión de la membrana viral y celular, seguida de la liberación de la cápside viral en el interior de la célula. (3) Retrotranscripción de una molécula de ARN de cadena sencilla a ADN de doble cadena. (4) Transporte del genoma viral al núcleo y su integración en el genoma celular. (5) Transcripción, replicación y transporte al citoplasma de los transcritos virales. (6) Traducción y procesamiento de las poliproteínas de los genes estructurales. (7) Ensamblaje de las partículas virales. (8) Gemación desde la célula de las partículas virales. (9) Maduración del virión. Figura cedida por la Dra. C. Reis.

1.2 Conceptos evolutivos.

1.2.1 Modelo de cuasiespecies.

Los virus cuyo material genético está compuesto por ARN, como es el caso del VIH, se constituyen naturalmente en distribuciones complejas de genomas íntimamente relacionados, aunque no idénticos, sometidos a un proceso continuo de variación genética, competición y selección (Figura 1.5) (Domingo y cols., 1985; Luring y Andino, 2010; Perales y cols., 2010). Este espectro de mutantes, conocido con el término de cuasiespecies (Eigen, 1996), se genera debido a las elevadas tasas de mutación de estos virus (medida en el VIH en $3,4 \times 10^{-5}$ sustituciones por sitio y por ciclo de replicación (Mansky y Temin, 1995)), por la falta de actividad correctora de sus polimerasas, y a su alta producción viral (de 10^8 a 10^9 viriones por día en el VIH) (Perelson y cols., 1996). Esta generación de variabilidad se ve aumentada por la existencia de recombinación (Iglesias-Sanchez y Lopez-Galindez, 2002; Jung y cols., 2002), que posibilita la regeneración de genomas libres de mutaciones deletéreas o de genomas con combinaciones de mutaciones. Esta inmensa variabilidad, es causa de que ninguna secuencia genómica definida caracterice a la población viral, sino que las distintas variantes conformen una secuencia consenso que es la media de todas las presentes en esta “nube de mutantes”. La distribución de variantes en estas cuasiespecies, como definió M.A. Nowak, sería el resultado de un proceso continuo de mutación-selección (Nowak, 2006).

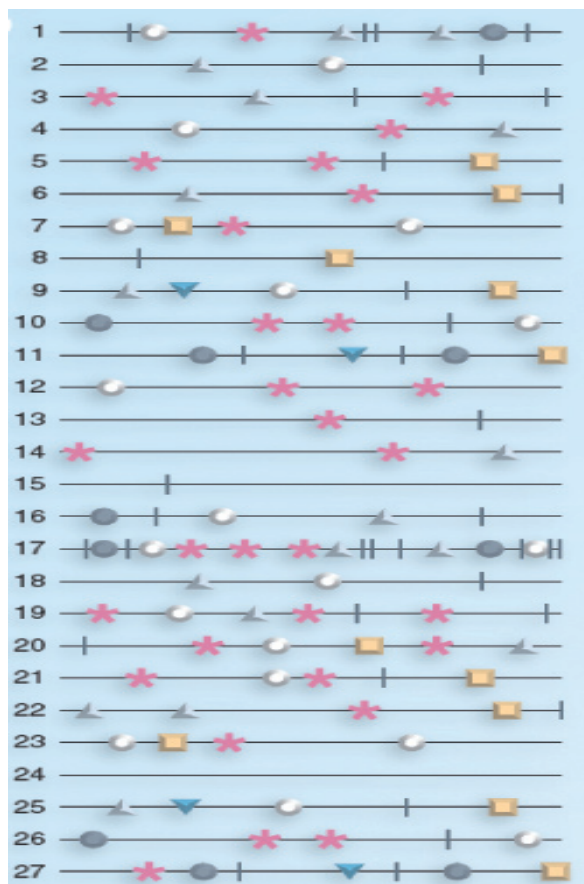


Figura 1.5 Distribución de una cuasiespecie viral. Representación esquemática de una distribución de mutantes que genera una cuasiespecie viral. Los diferentes genomas se representan por líneas horizontales y las mutaciones por diferentes símbolos sobre las líneas. Estas cuasiespecies presentan distintos genomas cada uno con un conjunto de mutaciones diferentes. Este espectro de mutantes es altamente dinámico y dispone de multitud de oportunidades para el cambio fenotípico. Figura modificada de Perales y cols 2010.

El término cuasiespecies fue introducido por M. Eigen y P. Schuster para describir teóricamente el tipo de estructura poblacional que podría haber mediado la auto-organización y la adaptabilidad de los replicones primitivos en las primeras fases de la generación de vida en la Tierra (Eigen, 1971).

El modelo describía el proceso de mutación-selección en poblaciones infinitas y con ambiente constante. La traslación de dicho modelo teórico a la dinámica de poblaciones virales, fue realizada por primera vez por E. Domingo en el bacteriófago Q β (Domingo y cols., 1978). Más tarde esta distribución poblacional se ha documentado en numerosos virus ARN, incluyendo importantes patógenos humanos y animales, como el VIH (Wain-Hobson, 1992), el virus de la gripe (Kuroda y cols., 2009; Lopez-Galindez y cols., 1985), el virus de la hepatitis humana C (VHC) (Martell y cols., 1992), el virus de la estomatitis vesicular (VSV) (Holland y cols., 1982; Novella, 2003) o el virus de la fiebre aftosa (VFA) (Domingo y cols., 1992; Sobrino y cols., 1983). A partir de esta traslación, los biólogos teóricos han encontrado en los virus ARN un modelo experimental para estudiar la dinámica de las cuasiespecies y los virólogos se han ayudado de este modelo teórico para explicar las propiedades observadas experimentalmente en las poblaciones virales.

Aunque el uso del término cuasiespecies no ha estado exento de polémica (Domingo, 2002; Holmes, 2010; Holmes y Moya, 2002; Wilke, 2005), mayoritariamente a causa de una mala interpretación del término, actualmente está ampliamente aceptado debido al extenso soporte experimental que ha recibido (Borderia y cols., 2010; Eigen, 1996; Luring y Andino, 2010; Ojosnegros y cols., 2010; Vignuzzi y cols., 2006). Esta distribución se ha observado “*in vivo*” mediante el análisis de clones moleculares aislados de poblaciones virales o utilizando las más recientes técnicas de ultra-secuenciación. Sin tener en cuenta el modelo de cuasiespecies virales, no se podrían explicar los cambios de eficacia biológica observados debidos a la composición del espectro de mutantes (Borderia y cols., 2010), la complementación e interferencia intra-poblacional (de la Torre y Holland, 1990; Perales y cols., 2007), la existencia de genomas memoria (Briones y cols., 2006; Ruiz-Jarabo y cols., 2000) o la inmensa capacidad adaptativa desplegada por estos virus.

Considerándose el espectro de mutantes generado por las poblaciones virales, se están explorando nuevas alternativas al tratamiento de enfermedades virales. Así, se han obtenido prometedores resultados con variantes con alta fidelidad de copia del virus de la polio, que podrían llevar a su utilización como virus vacunales, debido a su falta de capacidad adaptativa (Vignuzzi y cols., 2006; Vignuzzi, Wendt y Andino, 2008). En esta misma dirección, la extraordinaria variabilidad desplegada por el VIH, está teniéndose actualmente en cuenta para la generación de posibles nuevas vacunas, que generen respuestas mucho más amplias para contrarrestarla (Gaschen y cols., 2002).

Finalmente, el concepto de cuasiespecies ha superado la frontera de la virología y el comportamiento colectivo desplegado por este tipo de distribución poblacional se ha observado en colectividades celulares (Gonzalez-Garcia, Sole y Costa, 2002), priones (Li y cols., 2010) o ARNs catalíticos generados en estudios “*in vitro*” para estudiar el origen de la vida (Diaz Arenas y Lehman).

Todos estos datos demuestran el enorme impacto que el concepto de cuasiespecies ha tenido y tiene en el estudio de la evolución de poblaciones altamente variables, como los virus ARN, y la importancia que tendrá la consideración de esta distribución en el tratamiento de numerosas enfermedades ligadas a este tipo de poblaciones.

1.2.2 Eficacia Biológica.

Como enunció Darwin en su “Origen de las especies”, la supervivencia del más eficaz es la norma que dirige la evolución en una población compleja. La eficacia biológica es un concepto fundamental para el estudio de la evolución viral, que puede definirse como el parámetro que mide la adaptación replicativa de un organismo a su medio (Domingo y cols., 1999; Domingo y Holland, 1997). Es una variable relativa, ya que se mide con respecto a un virus de referencia por cultivos en competición, utilizándose generalmente un marcador fenotípico o genético que pueda diferenciar a las dos poblaciones virales (Holland y cols., 1991). En un ambiente constante, la eficacia biológica de un virus está determinada por las relaciones de las diferentes variantes dentro de las cuasiespecies, que llevarán a la imposición final de aquéllas que generen un mayor número de virus.

En una cuasiespecie, la distribución de secuencias nucleotídicas no es arbitraria sino que se centra en torno a una secuencia maestra, que es la de mayor eficacia biológica en esas condiciones ambientales, y que puede coincidir o no con la secuencia consenso de la población. La proporción de cada variante en la población global viene determinada por la eficacia biológica de cada una de ellas. En condiciones ambientales constantes, una cuasiespecie replicante se halla en equilibrio poblacional cuando mantiene su secuencia consenso invariable, aunque el espectro de mutantes sufra alteraciones por la generación estocástica de mutaciones. Este equilibrio puede alterarse por la generación de mutantes con mayor ventaja selectiva o por un cambio ambiental que confiera una ventaja selectiva a alguna variante hasta entonces minoritaria. Es decir, a menos que las condiciones ambientales permanezcan controladas y constantes, la superioridad o inferioridad de una variante en particular es transitoria. Durante las infecciones naturales, las perturbaciones del equilibrio poblacional ocurren frecuentemente, por ejemplo durante la transmisión, la aplicación de tratamiento o la aparición de mutantes, por lo que la estructura en forma de cuasiespecie proporciona a los virus ARN la capacidad de responder y adaptarse rápidamente a los cambios ambientales, asegurando así su supervivencia (Domingo y Holland, 1997).

1.2.3 Variabilidad.

La variabilidad que presentan las cuasiespecies virales es una consecuencia de las altas tasas de mutación y elevado recambio poblacional, que además constituye una característica esencial del virus, con importantes implicaciones en su comportamiento “in vivo”. Así, la composición y amplitud del espectro de mutantes son moldeados por la evolución y resultan de capital importancia en las características del virus, incluyendo la patogénesis.

La historia natural del VIH-1 condiciona la variabilidad desplegada por el virus en un hospedador. El tiempo de progresión a SIDA es muy variable entre individuos, teniendo una duración media entre 8 y 10 años post-infección en un individuo adulto sin tratamiento (Muñoz, Sabin y Phillips, 1997). De todos modos, el curso de la enfermedad normalmente sigue un patrón común y se divide en tres periodos diferenciados (Figura 1.6a). Tras la entrada del virus en el huésped, se inicia la fase aguda de la infección, en la que se produce un gran aumento de la carga viral (concentración de ARN viral en sangre), que conlleva una caída brusca del número de células CD4⁺. Esta fase es

seguida de un control de la replicación viral por el sistema inmune mediante las células $CD8^+$ y los anticuerpos neutralizantes, dando lugar a un período asintomático en el que el nivel de células $CD4^+$ se recupera y la carga viral decae. En dicha fase, que dura de media aproximadamente 8 años, el número de células $CD4^+$ cae progresivamente, mientras que el de $CD8^+$ aumenta. Finalmente, las células $CD4^+$ caen hasta alcanzar unos niveles irrecuperables, momento en el que la carga viral aumenta y las $CD8^+$ también disminuyen en número, entrando en la fase sintomática o de SIDA. Estos diferentes períodos producen cambios severos en el tamaño de población viral y, como consecuencia de ello, la diversidad que presentan las cuasiespecies dentro del individuo varía de manera considerable.

En numerosos estudios realizados en pacientes con VIH-1, se ha relacionado el aumento de la variabilidad viral con la progresión clínica (Anastassopoulou y Kostrikis, 2005; Essajee y cols., 2000; Ganeshan y cols., 1997; Wolinsky y cols., 1996). El estudio más consistente fue el realizado por el laboratorio de J. Mullins en el que se detectó un aumento de la variabilidad en el VIH-1 durante una infección en pacientes con progresión típica (véase Figura 1.6) (Shankarappa y cols., 1999).

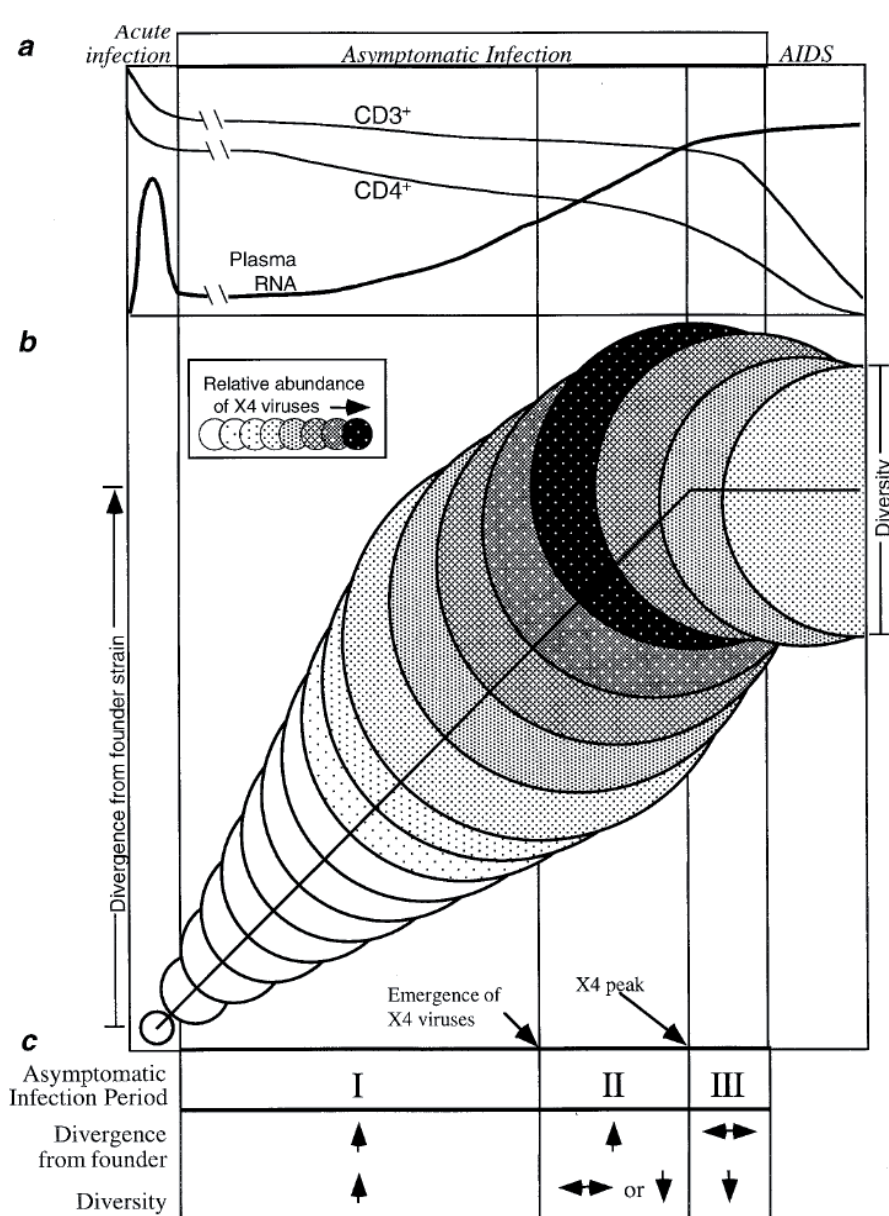


Figura 1.6 Descripción esquemática de los cambios en la divergencia y diversidad a lo largo de la infección por VIH-1 en pacientes típicos. A) Fases de la infección por VIH-1 desde el inicio hasta la aparición del SIDA. En esta figura se muestran los niveles de ARN viral en plasma, de linfocitos T $CD4^+$ y en general de las células $CD3^+$, que incluyen a los anteriores y a los linfocitos T $CD8^+$. B) Generación de divergencia y diversidad. La diversidad media de la población viral viene representada por el diámetro de los círculos. C) Tabla resumen de divergencia y diversidad según progresa la enfermedad. Los virus X4 son linfocito-tróficos, responsables de una progresión más rápida a SIDA. Figura tomada de Shankarappa y cols 1999.

En este trabajo se comprobó cómo a partir de una cuasiespecie que era homogénea en el momento de la transmisión, ésta iba aumentando su diversidad (número de diferencias entre los miembros de la población en un momento dado) a lo largo de la infección viral para estabilizarse justo antes de entrar en fase de SIDA. Esto mismo se observó para la divergencia (número de sustituciones acumuladas durante el proceso evolutivo con respecto al virus inicial). Estos resultados sugerían que la evolución del VIH seguía un modelo consistente en el que el virus evadía las respuestas del sistema inmune gracias a esta gran variabilidad desplegada. La complejidad de las cuasiespecies virales iba aumentando a medida que el virus escapaba de las sucesivas barreras defensivas impuestas por el huésped, y la enfermedad progresaba. Una vez exhausto el sistema inmune, éste no ejercía tanta presión selectiva y la variabilidad se estabilizaba e incluso disminuía, al no necesitar el virus escapar más, desarrollándose la fase de SIDA (Quinones-Mateu y Arts, 2006). No sólo se observó esta relación entre progresión clínica y variabilidad, sino que también se vio en pacientes una relación directa entre aumento de variabilidad y de la eficacia biológica (Troyer y cols., 2005).

En esta misma dirección, varios estudios han descrito la tendencia a la transmisión, debido a su mayor eficacia, de variantes virales con regiones variables (V1-V4) de la envuelta más cortas (Derdeyn y cols., 2004). Sin embargo, estos virus con las regiones variables cortas, son más fácilmente neutralizables, por lo que durante el transcurso de la infección aparecen variantes virales con regiones más largas que pueden evadir la respuesta inmune.

En definitiva, una alta variabilidad permite al virus adaptarse con extraordinaria facilidad a los ambientes cambiantes, por lo que puede evadir cualquier tipo de respuesta inmune en la mayoría de los casos, y, por lo tanto, está relacionada con el aumento de eficacia biológica de los mismos.

1.2.4 Tamaño poblacional.

Durante la infección natural, el VIH sufre cambios radicales de tamaño poblacional debido a las diferentes fases que atraviesa la infección. El virus sufre drásticos cuellos de botella en la primoinfección, que en un 60% de los casos está causada por una única variante (Salazar-Gonzalez y cols., 2008). También se producen importantes cuellos de botella cuando el virus escapa de la respuesta inmune, al desarrollar resistencia al tratamiento o debido al efecto fundador que se produce al colonizar algún órgano. Por otro lado, se producen crecimientos poblacionales muy grandes tras estos cuellos de botella como en la fase aguda de la infección o en los picos de viremia cuando el sistema inmune o el tratamiento fallan. Estos cambios de tamaño poblacional, a su vez, conllevan cambios en la complejidad de la población contribuyendo en gran parte a las características y evolución de la infección viral.

La tasa a la que una mutación se fija en la población debido a fuerzas deterministas o estocásticas depende del tamaño de población efectiva (N_e), que se puede definir como el tamaño de una población idealizada que se relaciona aleatoriamente y tiene la misma frecuencia de cambios que la población que se está estudiando (Vandamme, 2009). Este tamaño de población efectiva es uno de los determinantes principales en la frecuencia de mutaciones a lo largo del tiempo. Así, las tasas de evolución se ven muy influenciadas por generaciones con menores tamaños de población efectiva, como ocurre en el caso de los cuellos de botella genéticos. En estos casos, la mayoría de los

cambios se fijan debido a efectos aleatorios, produciéndose lo que se denomina deriva genética. Por otro lado, cuando se dispone de tamaños poblacionales efectivos grandes, las mutaciones se fijan en la población exclusivamente gracias a su eficacia biológica en un ambiente particular dado. Por lo tanto, cuanto menor es el tamaño de población efectiva, mayor es el efecto de eventos aleatorios y de la deriva genética, y menor es la importancia de las fuerzas selectivas.

En este contexto, surgió un modelo conocido como “Trinquete de Muller”, que proponía un proceso en el que se produce una acumulación y fijación de mutaciones deletéreas en poblaciones de organismos asexuales con un bajo N_e (Muller, 1964). En dicho proceso, la acumulación de dichas mutaciones sería irreversible debido al bajo tamaño poblacional, que impediría procesos de recombinación y limitaría las posibilidades de generación de variabilidad del virus y de compensación. Consecuentemente, la predominancia de deriva genética que ocurre en esta situación, podría llevar al virus a pérdidas de eficacia biológica. Si este bajo N_e se mantiene a lo largo del tiempo, por ejemplo con pases seriados placa a placa, se puede llegar a la extinción de la población viral como fue descrito por (Domingo y cols., 2005; Lynch y cols., 1993). En el caso del VIH-1 el Trinquete de Muller es operativo y se ha demostrado en estudios “in vitro” (Yuste y cols., 1999). Asimismo, la acción del trinquete se ha estudiado en otros virus de ARN, como el bacteriófago $\Phi 6$ (Chao, 1990), el VSV (Clarke y cols., 1993; Duarte y cols., 1992) y el VFA (Escarmis y cols., 1996).

En otros sistemas biológicos, el efecto del trinquete de Muller también está ampliamente aceptado, al haber sometido a repetidos cuellos de botella a diversas poblaciones experimentales, como por ejemplo: *Drosophila melanogaster* (Mukai y cols., 1972), *Caenorhabditis elegans* (Keightley y Caballero, 1997), *E. coli* (Kibota y Lynch, 1996), o *Aspergillus nidulans* (Bruggeman y cols., 2003).

El estudio de los cuellos de botella poblacionales y el proceso de recuperación de eficacia posterior es de vital importancia para la comprensión de las infecciones virales. Esto es debido a que la mayoría de las transmisiones virales implican un paso por cuello de botella poblacional. A partir de este momento, los virus comienzan a replicar a altas tasas y a generar grandes tamaños poblacionales.

1.2.5 Selección natural.

La adaptación mediante la selección natural es el proceso más importante en Biología. No se puede entender la evolución a lo largo del tiempo de los diferentes organismos sin tener en cuenta este concepto, “nada tiene sentido en biología si no es a la luz de la evolución” (Dobzhansky, 1964). Por lo tanto, resulta esencial en los estudios evolutivos entender las fuerzas selectivas operando en cada momento. Las diferentes fuerzas evolutivas que pueden afectar a las poblaciones son: las mutaciones, la selección natural y la antes mencionada deriva genética. La mutación, aunque es la fuente de la variación genética, por sí misma es la fuerza evolutiva más débil. Una sola mutación no puede cambiar su frecuencia en la población, nada más que apareciendo de forma independiente una y otra vez. La deriva genética, como ya se vio anteriormente, es la principal fuerza evolutiva cuando el tamaño poblacional es muy bajo. Por último, la principal fuerza evolutiva operando en poblaciones de gran tamaño es la selección natural.

En la selección natural, las mutaciones se fijan o desaparecen de la población en función de su eficacia biológica y del efecto que sobre esta variable tienen en la población. Se distinguen tres tipos de mutaciones: las mutaciones beneficiosas, que suponen ganancia de eficacia, mutaciones deletéreas, que tienen efectos negativos sobre la eficacia, y mutaciones neutrales, aquéllas que no tienen ningún efecto en la eficacia. El efecto de la selección natural sobre las mutaciones también hace distinguir entre dos tipos de selección: la selección positiva sería aquélla cuyo efecto es incrementar la frecuencia de mutaciones beneficiosas, mientras que la selección negativa o purificadora tendería a eliminar las mutaciones deletéreas. Finalmente, las mutaciones neutrales serían predominantes en procesos de deriva genética, aunque existe otra corriente teórica, la teoría neutralista, que establece que la gran mayoría de los cambios evolutivos son debidos a la deriva genética de mutantes selectivamente neutros (Kimura, 1989). Sin embargo, la aplicación simple de la teoría neutralista para poblaciones virales de gran tamaño no parece apropiada, debido a que varias de las asunciones de la teoría chocan con las características desplegadas por virus como el VIH (Rouzine, Rodrigo y Coffin, 2001).

Para el estudio de la selección natural, se hace uso de la huella molecular dejada por dicha selección. Se pueden utilizar las secuencias nucleotídicas para estimar el tipo de selección a la que han sido sometidas. Aunque se utilizan diversos métodos, los más ampliamente utilizados, tienen en cuenta el tipo de mutaciones encontradas en las secuencias a estudiar. Así, se examina si los cambios observados en las secuencias de nucleótidos conllevan o no cambios aminoacídicos. Las mutaciones sinónimas serían aquéllas que no producen un cambio de aminoácido debido a la degeneración del código genético, entre tanto que las no sinónimas implicarían la codificación de un aminoácido diferente en la proteína correspondiente. Una vez examinados los diferentes tipos de mutaciones, se puede asumir que la dominancia de uno u otro tipo indica la actuación de diferentes pautas de selección. La preponderancia de mutaciones sinónimas indicaría la actuación de la presión negativa, ya que los únicos cambios que se observarían serían los permitidos por la selección, que habría eliminado mayoritariamente los no sinónimos. Por el contrario, una presencia mayoritaria de mutaciones no sinónimas sería indicativa de presión positiva, ya que los cambios que producen alteración en las proteínas habrían sido seleccionados por sus efectos beneficiosos en la población. Teniendo en cuenta estos conceptos se puede estimar el tipo de fuerzas selectivas que han operado a lo largo de la evolución de las secuencias estudiadas.

1.2.6 Paisaje de eficacias.

Fue el biólogo teórico John Maynard Smith, el que por primera vez introdujo el concepto de “espacio de secuencias” al proyectar todas las proteínas de una determinada longitud, de tal manera que los vecinos más próximos difiriesen únicamente en un aminoácido (Maynard-Smith, 1970). Varios años antes (1932), el genético de poblaciones Sewall Wright ideó una atractiva manera de escenificar el cambio en las frecuencias de alelos, dibujando la eficacia de estos alelos en función de su frecuencia (Wright, 1932). Esta aproximación dio lugar a paisajes adaptativos, donde los picos correspondían a alelos con mayor frecuencia en la población, mientras que en los valles recaían la variantes con frecuencias bajas. La idea de juntar el espacio de secuencias con estos paisajes adaptativos surgió de la colaboración entre Manfred Eigen y Peter Schuster, creando el concepto de paisaje de

eficacias (Eigen y Biebricher, 1988). En este tipo de representación se traza un mapa relacionando el genotipo y la eficacia de cada secuencia. El paisaje estaría formado por montañas y valles, que representarían la eficacia biológica, siendo las montañas o picos adaptativos valores elevados de eficacia y los valles, valores bajos de eficacia. Según esto, los distintos genotipos se distribuirían a lo largo de este paisaje, tendiendo a agruparse los genotipos similares y a separarse los genotipos no relacionados. La rugosidad o complejidad de este paisaje, vendrá determinada por la magnitud de la epistasia del sistema, definida como la interacción entre diferentes partes del genoma para determinar una característica fenotípica concreta (Hinkley y cols., 2011). Este tipo de representación ha sido ampliamente utilizado para describir y estudiar los procesos evolutivos de diferentes organismos, en especial de poblaciones virales (Elena y Sanjuan, 2003).

Sin embargo, los paisajes de eficacias utilizados en diferentes estudios han sido habitualmente teóricos o estimados, porque normalmente no se han utilizado datos experimentales de eficacia. Se han establecido paisajes de eficacias teóricos (Biebricher y Eigen, 2005; Feng y cols., 2007) y estudiado “in silico” la evolución de secuencias generadas al azar (Li y Wilke, 2004; Tejero, Marin y Montero, 2011). Además, se han representado paisajes teóricos para explicar procesos evolutivos (Burch y Chao, 2000; Eigen y Biebricher, 1988; Elena y Sanjuan, 2003). En cuanto a paisajes generados con datos experimentales, se han utilizado valores de eficacia de diferentes variantes virales (Deforche y cols., 2008; Fernandez, Clotet y Martinez, 2007) u otro tipo de organismos y moléculas (Hayashi y cols., 2006; Pitt y Ferre-D’Amare, 2010), utilizando diversos métodos, y se ha representado el paisaje generado por sus secuencias.

1.2.7 Redes Neuronales Artificiales.

En la presente tesis se ha investigado la utilización de Redes Neuronales Artificiales (siglas en inglés ANN), para la representación de paisajes de eficacias experimentales con capacidad predictiva. Estas ANN son instrumentos computacionales ampliamente utilizados para realizar estadísticas no lineales. Las ANN han sido utilizadas en diferentes campos para realizar agrupamientos, clasificaciones, reconocimiento de patrones, búsqueda de datos, etc... Antes de que pueda desempeñar la función deseada, toda ANN necesita ser entrenada. El proceso de entrenamiento consiste en la presentación de series de vectores de datos para que la ANN cambie sus valores internos (vectores de referencia o sinápticos). Las ANN más frecuentemente utilizadas en Biología son aquellas con un aprendizaje sin supervisión, donde los valores por los que se clasifica cada muestra no necesitan ser determinados *a priori*.

Los mapas auto-organizativos (SOM), desarrollados por Kohonen (Kohonen, 2001), son un tipo de redes neuronales no supervisadas que llevan a cabo la clasificación de los datos de tal modo que generan un mapa bidimensional, en el que los datos se ordenan topológicamente (Figura 1.7). El modelo de red neuronal de estos SOM de Kohonen proyecta espacios de entrada de muchas dimensiones en una estructura topológica bidimensional. Este método de mapeo está determinado por el algoritmo SOM, que está definido como un proceso no supervisado y paramétrico, es decir no necesita ningún valor de clasificación de los vectores de entrada ni hace asunciones sobre la distribución de los datos, y consiste en un conjunto de vectores N-dimensionales organizados en

una matriz discreta bidimensional. Cuando una red SOM es creada para ser entrenada, todos sus vectores, los denominados vectores sinápticos, se inician con valores aleatorios bajos. Durante la fase de aprendizaje, un vector de entrada es seleccionado del conjunto de entrenamiento, se computa el vector sináptico de la red más cercano (unidad que mejor se aproxima, siglas en inglés bmu) y este vector y todos los de su área próxima se cambian para aproximarse al vector de entrada. Este proceso es repetido de manera continua para todo el conjunto de vectores de entrenamiento. El algoritmo SOM hace que el mapa se auto-organice, de tal manera que los nodos próximos de salida respondan a vectores similares de entrada. El mapa SOM resultante agrupa los vectores de entrenamiento de entrada y puede ser utilizado para localizar vectores de entrada desconocidos y descubrir su similitud con aquellos que conforman el conjunto de entrenamiento. Más aun, si la red es sometida a un proceso de etiquetado en el que a cada nodo se le asigna una medida asociada, dicha red se puede utilizar para estimar esta medida para vectores de entrada no presentes en el conjunto de entrenamiento.

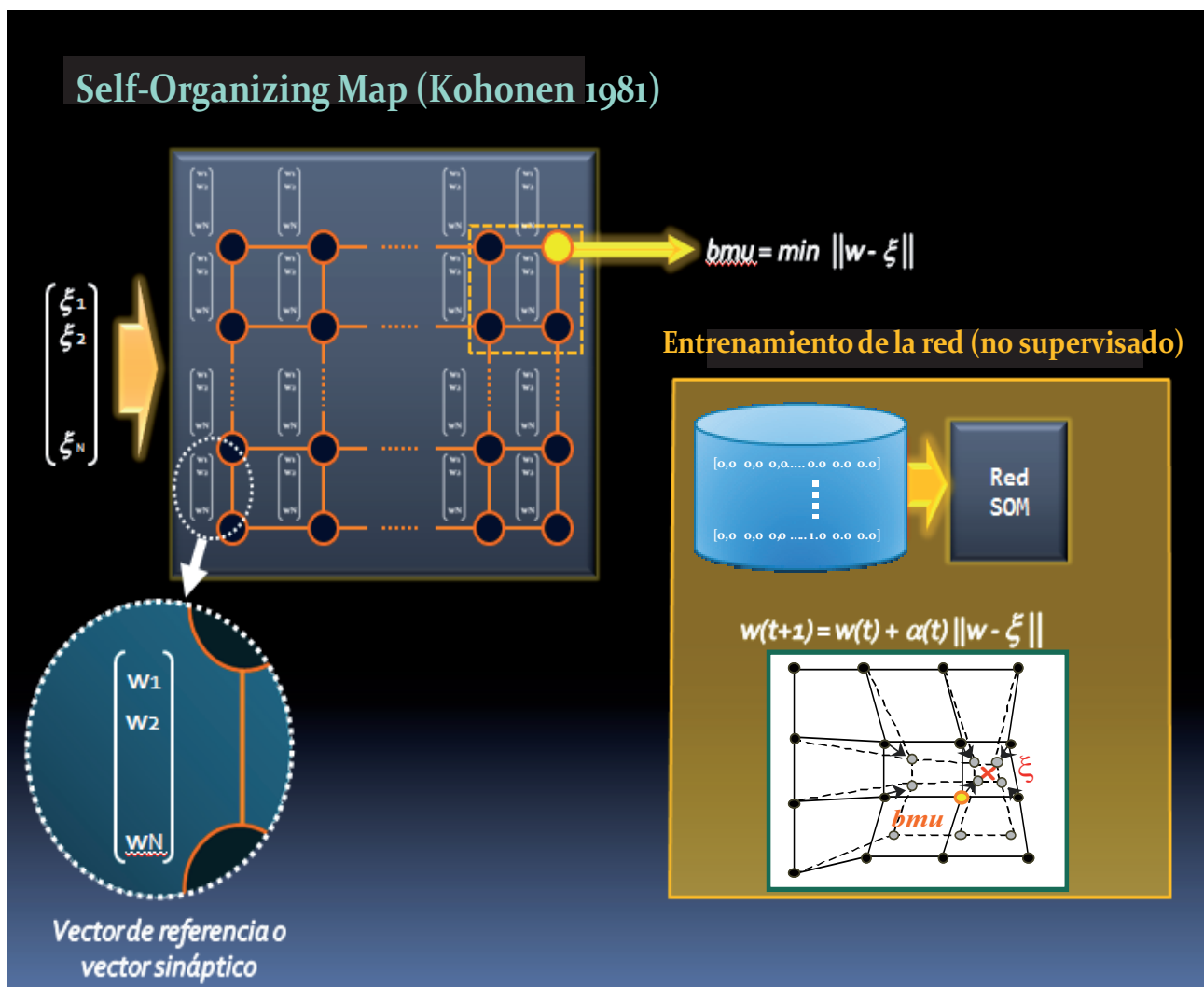


Figura 1.7 Esquema del funcionamiento de los SOM. A una red SOM compuesta por vectores sinápticos (w), se le presenta un conjunto de vectores de entrada (ξ) de los que se halla el vector sináptico más cercano a cada uno de ellos (bmu). En la fase de aprendizaje, este bmu y todos los de su entorno se cambian para aproximarse al vector de entrada mediante el algoritmo de entrenamiento indicado en la figura. Este proceso de entrenamiento no supervisado continúa hasta que se consigue una clasificación óptima de los vectores de entrada. Figura cedida por la Prof. S. Delgado

Para determinar el vector sináptico más cercano a un vector de entrada, SOM utiliza distancias Euclídeas, lo que representa un problema cuando se utilizan vectores de datos no numéricos, como es el caso de las secuencias nucleotídicas. En estos casos, para procesar las secuencias utilizando redes SOM, es necesario transformarlas en vectores numéricos. Para solventar ese problema el grupo del Prof. Morán en la Universidad Complutense de Madrid junto con el Dr. Briones del Centro de Astrobiología y la Prof. Delgado de la Universidad Politécnica de Madrid han desarrollado una codificación basada en los vértices de un tetraedro, que será utilizada en la presente tesis (Delgado, 2009).

Gracias a esta metodología, se puede abordar la tarea de organizar secuencias nucleotídicas topológicamente de acuerdo a sus similitudes e intentar generar paisajes de eficacias, asociando los valores de eficacia biológica conocidos de los genomas estudiados. Una vez obtenido este paisaje por redes neuronales, éste se utilizaría para estimar eficacias desconocidas de secuencias relacionadas, constituyendo una posible herramienta muy útil en estudios evolutivos.

1.3 Antecedentes.

A lo largo de varios años en el servicio de Virología Molecular del Centro Nacional de Microbiología se vienen estudiando los procesos que ocurren “in vitro” al someter al VIH-1 a cambios de tamaño de población, simulando los cambios producidos en la infección real (Sanchez-Palomino y cols., 1993; Yuste y cols., 2005; Yuste, Lopez-Galindez y Domingo, 2000; Yuste y cols., 1999). Así, se sometió al virus a pases seriados placa a placa para simular los cuellos de botella que sufre, por ejemplo en la transmisión, al actuar la respuesta inmune o al aplicar tratamientos. Por otro lado, se han realizado pases en gran tamaño poblacional, como ocurre en la infección aguda. El objetivo de estos experimentos es estudiar la variación que se produce en la eficacia biológica y los cambios asociados a ella. Para estudiar las poblaciones virales, se utiliza la eficacia al ser el mejor parámetro que mide la adaptación replicativa de un organismo a su medio.

En la primera fase del estudio, se sometió a 10 clones biológicos, provenientes de un mismo virus (S61), a 15 pases seriados placa a placa. El objetivo de este estudio era comprobar si estaba operando el Trinquete de Muller en el VIH-1; modelo que predice la acumulación irreversible de mutaciones deletéreas al someter al virus a cuellos de botella seriados que puede llevar a la extinción del virus (véase Apartado 1.2.4). Los resultados obtenidos fueron una drástica disminución de la eficacia biológica de los virus en muy pocos pases, incluyendo virus que no fueron capaces de completar los 15 pases (Yuste y cols., 1999). Este proceso estaba mediado por la acumulación de pocas mutaciones (Yuste, Lopez-Galindez y Domingo, 2000).

En la segunda fase, los virus obtenidos tras los pases en “cuello de botella” fueron pasados seriadamente en gran tamaño poblacional. Con ello, se quería estudiar si estos virus recuperaban la eficacia biológica y, si lo hacían, qué cambios estaban asociados a esta recuperación.

En los primeros 10 pases, se pudo observar un aumento significativo en la eficacia biológica global de todos los virus; sin embargo había virus con incrementos significativos de eficacia, junto con otros en los que no se producía ningún cambio de eficacia (Yuste y cols., 2005).

Para comprobar si el incremento de eficacia era debido a un aumento de producción viral, se midió el título y la producción de proteína p24 a pase 1 y pase 11. Los resultados obtenidos fueron un aumento significativo del título viral sin diferencias en los niveles de proteína p24, lo que indicaba un aumento en la eficiencia viral sin aumento en la producción.

Una vez obtenidos estos resultados, se procedió a estudiar las mutaciones que se fijaban en la secuencia global asociadas a esta recuperación de eficacia. Sorprendentemente, las mutaciones fijadas eran muy pocas, tan solo 25 entre todos los virus, y únicamente se producían en 7 de los 12 virus estudiados. No sólo eran pocas las mutaciones, sino que había virus que aumentaban su eficacia sin fijar ninguna mutación, mientras que otros con mutaciones fijadas no sufrían ningún cambio (Borderia y cols., 2010).

Con estos antecedentes, se inició la presente tesis doctoral, en la que se pretende comprender el proceso por el cual el VIH-1 recupera eficacia y estudiar la evolución del virus en dicho proceso de recuperación llevado a cabo “in vitro”. En este estudio se examinará principalmente el papel de la heterogeneidad genética de las cuasiespecies de los virus estudiados en la recuperación de eficacia, así como el papel de ciertas mutaciones observadas a lo largo de todo el proceso.

2. Objetivos

Esta tesis plantea el estudio del proceso de recuperación “in vitro” de la eficacia biológica de clones biológicos del VIH-1 por pases seriados en gran tamaño poblacional. Para ello, se han realizado diversos abordajes: genéticos, fenotípicos, moleculares y teóricos, con el fin de tener un conocimiento lo más extenso posible de la evolución “in vitro” del VIH-1.

Los objetivos concretos son los siguientes:

- 1) Análisis de la recuperación de eficacia biológica de 10 clones biológicos del VIH-1, previamente debilitados debido a sucesivos cuellos de botella poblacionales, durante 30 pases seriados “in vitro” en gran tamaño poblacional.
- 2) Determinación de las mutaciones aparecidas y su dinámica durante los pases en gran tamaño poblacional. Para ello, se compararán las secuencias nucleotídicas de los clones virales después de 20 y 30 pases seriados con las secuencias anteriores a los pases (Yuste y cols., 2005).
- 3) Estudiar las consecuencias biológicas de distintas mutaciones aparecidas durante la pérdida y recuperación de eficacia biológica por mutagénesis dirigida.
- 4) Analizar el papel de la heterogeneidad de las cuasiespecies virales en la recuperación de la eficacia biológica.
- 5) Reconstruir la evolución y el paisaje adaptativo que describen estos virus en su proceso de recuperación de eficacia, mediante abordajes filogenéticos, moleculares y computacionales.

3. Materiales y Métodos

3.1 Cultivos celulares.

La línea celular linfoide CD4+ denominada MT-4 (Harada, Koyanagi y Yamamoto, 1985) ha sido la principal línea utilizada en la presente tesis. Se utilizó en los pases seriados de recuperación del virus, así como en los ensayos de competición para hallar la eficacia biológica del virus. Además, se utilizaron otras líneas celulares para diversos experimentos, como las células MT-2 (Harada, Koyanagi y Yamamoto, 1985) para las titulaciones de los virus y la línea 293T (DuBridge y cols., 1987) para transfecciones de distintos plásmidos recombinantes portadores de las mutaciones estudiadas en esta tesis.

Todas las líneas celulares, excepto la 293T, se cultivaron en medio RPMI 1640 (Cambrex) suplementado con suero bovino fetal (SBF) al 10% (PAN Biotech), L-glutamina 2 mM, penicilina 100 unidades (U)/ml, estreptomicina 100 µg/ml (Cambrex). A este medio suplementado se le denominará a partir de ahora medio completo. La línea celular 293T se cultivó en medio DMEM (Lonza) suplementado de igual manera. Todos los cultivos celulares e infecciones se mantuvieron a una temperatura de 37°C en atmósfera de CO₂ al 5% con una humedad relativa del 95%.

La viabilidad de los distintos cultivos celulares fue determinada por cuantificación en el microscopio de la proporción de células vivas y muertas en una cámara de Neubauer por exclusión de azul tripan (Sigma).

3.2 Virus e infecciones.

3.2.1 Virus s61.

El virus inicial se aisló por co-cultivo de linfocitos de sangre periférica del paciente 61/89 en el año 1989 con linfocitos de donante sano en presencia de interleukina 2 (IL-2) en el Servicio de Diagnóstico y Referencia de Retrovirus (Centro Nacional de Microbiología, Instituto de Salud Carlos III). Posteriormente, esta semilla viral inicial se usó para infectar células MT-4, obteniéndose de esta manera la semilla viral denominada s61 (Sanchez-Palomino y cols., 1993).

3.2.2 Descripción de los clones virales D15, E11, G7, H12, I15, K15.

Estos clones virales son clones biológicos obtenidos a partir del virus s61 por plaqueo en células MT-4, y posteriormente sometidos a cultivo placa a placa durante 15 pases (Yuste, Lopez-Galindez y Domingo, 2000; Yuste y cols., 1999). Todos estos clones vieron su eficacia biológica disminuida tras los pases placa a placa. Algunos de ellos se extinguieron antes de los 15 pases, como es el caso de los clones E, G y H en los que sólo se pudo realizar pases placa a placa hasta los pases 7, 12 y 13 respectivamente. De cada uno de estos clones se disponía de una serie de placas resuspendidas en 200 µl de medio completo, aisladas en el último plaqueo que se pudo realizar (Yuste y cols., 1999).

Para llevar a cabo pases de recuperación de eficacia biológica, se seleccionaron 2 placas de cada clon viral a excepción de los clones E11 y H12, de los cuales sólo se seleccionó una, y se obtuvo una semilla inicial infectando con 100 µl de sobrenadante de cada placa 1×10^5 células MT-4. Se disponía, por lo tanto, de 6 linajes virales distintos (D, E, G, H, I, y K), y dos réplicas para cada uno de ellos a excepción de los linajes E y H (Figura 3.1).

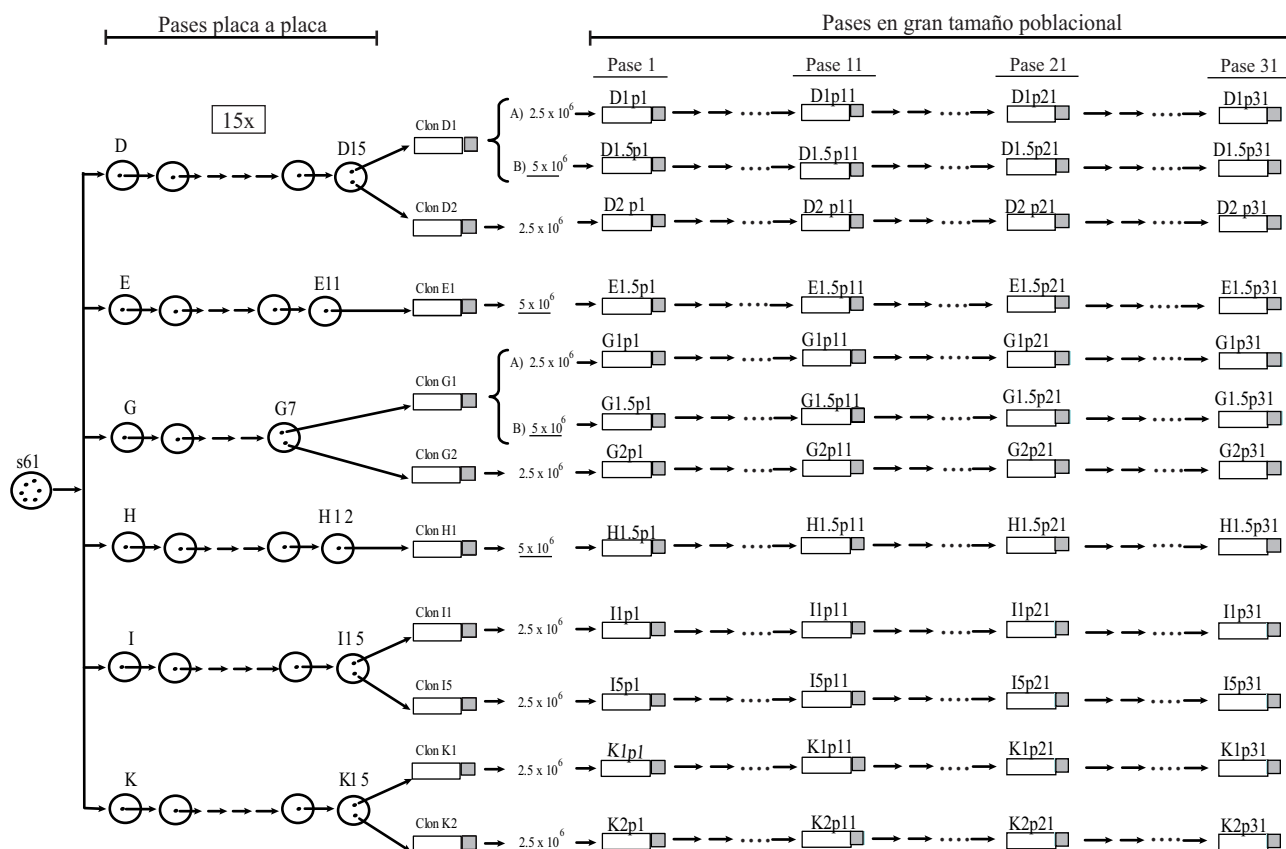


Figura 3.1 Genealogía de los virus estudiados. Representación de los pases seriados placa a placa (círculos en la parte izquierda de la figura) llevados a cabo en trabajos previos (Yuste y cols 1999). Se obtuvieron uno o dos clones de las poblaciones finales tras los cuellos de botella seriados y se designaron como D1, D2, E1, G1, G2, H1, I1, I5, K1 y K2. Con estos clones se realizaron pases en gran tamaño poblacional, indicados por flechas y puntos (representando los rectángulos las botellas de cultivo), en $2,5 \times 10^6$ y 5×10^6 células MT-4. Las poblaciones virales se indican con letras correspondientes al clon utilizado, seguido de p1 para la población inicial, p11 para el pase 11, p21 para el pase 21 y p31 para el pase 31. Para los clones D1, E1, G1 y H1, el número 5 después del punto indica que estos clones se pasaron en 5×10^6 células MT-4. Los clones D1 y G1, que están agrupados por llaves, fueron pasados en paralelo en $2,5 \times 10^6$ (designado A) y 5×10^6 células (designado B).

3.2.3 Pases del virus en gran tamaño poblacional.

Después de los pases placa a placa, se llevaron a cabo pases adaptativos de poblaciones virales en gran tamaño para recuperar la eficacia biológica de estos virus debilitados. Para ello, se infectaron $2,5 \times 10^6$ ó 5×10^6 células MT-4 con los clones obtenidos tras los pases placa a placa, a una multiplicidad de infección (m.o.i.) de 0,01 unidades formadoras de placa (UFP)/célula (Figura 3.1). El virus fue recuperado al recoger el sobrenadante de los cultivos cuando el efecto citopático era completo (de 5 a 7 días después de la infección). En cada pase, se infectaron células MT-4 con una dilución 1/10 del sobrenadante recuperado del pase anterior en el mismo volumen que el inóculo inicial, teniendo en cuenta que los títulos virales en VIH-1 no cambian significativamente durante los pases seriados (Sánchez-Palomino, 1993). Se mantuvieron en todo momento cultivos sin infectar en paralelo como control de la viabilidad celular y de contaminaciones.

3.3 Eficacia biológica de los virus.

3.3.1 Ensayo de eficacia biológica.

La determinación de la eficacia biológica se llevó a cabo mediante ensayos de competición (Holland y cols., 1991). El ensayo consiste en la co-infección de cultivos celulares con el virus a ensayar y un virus de referencia (J1). Este virus J1 fue elegido porque tiene una delección de 45 nucleótidos en la región V1-V2 del gen *env*, que permite diferenciarlo del resto de los virus. Estas co-infecciones se llevaron a cabo a diferentes proporciones iniciales entre el virus a estudiar y el de referencia (J1) (1:9, 1:1 y 9:1), a una m.o.i. de 0,01, y se compitieron durante 5 pases. En cada pase, se utilizó una dilución 1/10 del sobrenadante del pase anterior para infectar 1×10^5 células MT-4. La proporción de los dos virus se cuantificó por un análisis de GeneScan. La proporción del variante estudiado con respecto al virus de referencia (R_n) es dividido por su proporción en el pase inicial (R_o), y el Ln de este valor (R_n/R_o) es representado gráficamente frente a los pases de competición para obtener el vector de eficacia. Los valores de eficacia se obtienen de la pendiente exponencial de cada vector (Yuste y cols., 1999). Las diferencias entre los valores de eficacia de cada virus con respecto a su pase inicial, se examinaron comparando las pendientes de los vectores obtenidos en sus respectivas competiciones con el método implementado en el programa GraphPad.

3.3.2 Análisis por “GeneScan”.

Se realizó una extracción de ADN por el método de lisis directa, ya que en estos experimentos partíamos de un número muy bajo de células (1×10^5). Esta extracción consiste en el tratamiento de las células infectadas con un buffer con detergentes conteniendo Tris pH 8,3 (1 M), KCl (1 M), $MgCl_2$ (1 M), gelatina (10 mg/ml), NP40 (0,45%) y Tween 20 (0,45%). Tras añadir a este buffer proteinasa K (10 mg/ml), se incubaron las células 1 hora a 58°C y 10 minutos a 98°C, obteniéndose de esta manera el ADN celular. El ADN de la región V1-V2 de los virus en competición se amplificó por la reacción en cadena de la polimerasa (en inglés, PCR) utilizando uno de los cebadores marcado con fluoróforo 6-FAM. Los cebadores utilizados fueron: el 34-Sentido marcado, 5'-6FAM-GTCACAGTCTATTATG GGGTACCTGTGT-3'; y el 31-Antisentido, 5'-ACCTCAGTCATTACACAGGCCTG CAGCGC-3'. A continuación, se mezcló 1 µl del producto de PCR con 9 µl del marcador de tamaño molecular ROX 1000 size standard (Applied Biosystems) diluido 1/10 en formamida. Las muestras fueron sometidas a electroforesis capilar en un secuenciador ABI PRISM 3700 y los datos fueron analizados utilizando el programa GeneMapper v 4.0. Las áreas de los picos obtenidas por este método, correspondientes al ADN presente en los cultivos procedente del virus analizado y del virus de referencia, fueron utilizadas para cuantificar la cantidad de cada virus en la competición y así, calcular la eficacia de cada virus con respecto al estándar J1.

3.4 Caracterización genética de los virus.

3.4.1 Obtención de las secuencias nucleotídicas.

3.4.1.1 Extracción del ADN de las células de los cultivos infectados.

En el caso de los pases en gran tamaño poblacional, como se partía de $2,5$ o de 5×10^6 células infectadas con el virus, se siguió un método de extracción estándar con fenol:cloroformo (Perucho y cols., 1981). En primer lugar, se lavaron las células con tampón fosfato salino 1X, resuspendiéndose posteriormente en 1 volumen de solución de lisis A (Tris 10 mM pH 7,5, ácido etilendiaminotetraacético (EDTA) 10 mM, NaCl 0,15 M) a los que se añade proteinasa K, a una concentración final de 200 μ g/ml, y dodecil sulfato sódico (SDS) al 10%. Tras 1 hora de incubación a 65°C las muestras se dejaron toda la noche a 37°C, se añadió 1 volumen de solución de lisis B (Tris 10mM pH 8, EDTA 10 mM; NaCl 0,65 M) y se realizaron tres extracciones con fenol:cloroformo:alcohol isoamílico (25:24:1), seguidas de una extracción con cloroformo:alcohol isoamílico (24:1). El ADN contenido en la fase acuosa se precipitó con 2 volúmenes de etanol absoluto. Posteriormente, el ADN se resuspendió en agua estéril para su uso posterior. Se determinó la concentración de ADN mediante espectrofotometría, haciendo uso para ello de un NanoDrop (Thermo Scientific). El ADN se almacenó a una temperatura de -20°C.

3.4.1.2 Amplificación de las secuencias de los genomas virales.

Para la obtención de las secuencias correspondientes a los genomas provirales integrados, se partió del ADN extraído de los cultivos celulares para llevar a cabo las amplificaciones por PCR. Previamente, se seleccionaron distintas parejas de cebadores disponibles en el laboratorio para poder realizar la amplificación del genoma viral completo en 3 fragmentos (Figura 3.2) (Tabla 3.1). Asimismo, se seleccionaron también unos cebadores internos a los anteriores para, en caso necesario, realizar amplificaciones anidadas.

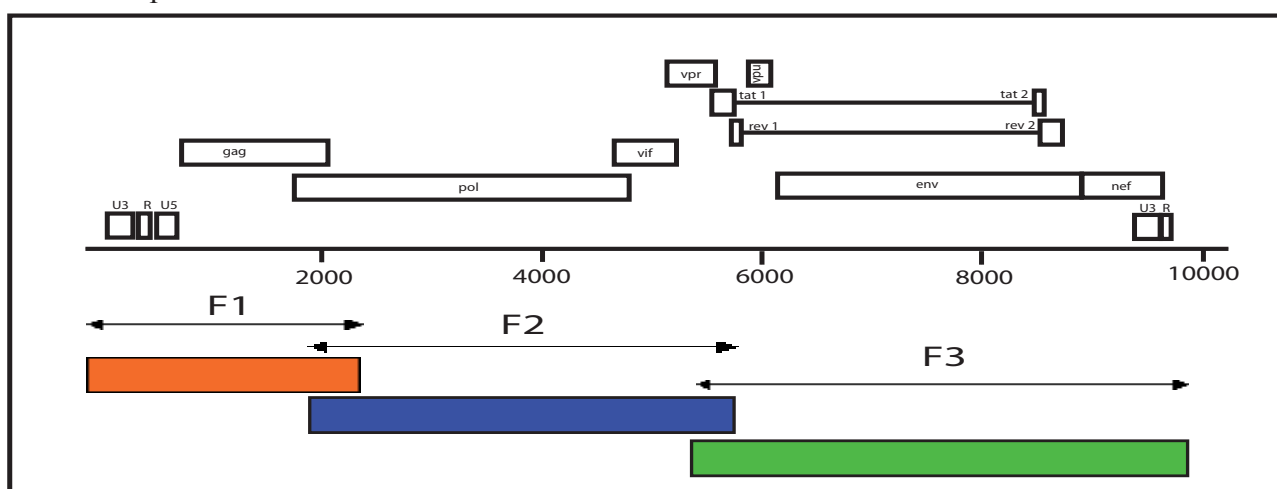


Figura 3.2 Esquema del proceso de amplificación del genoma completo. Se indican los tres fragmentos en los que se amplifica el genoma completo del virus para su posterior secuenciación. El fragmento 1 (F1) abarca desde la LTR en 5' hasta la zona de la RT en *pol*, el fragmento 2 (F2) va desde la p7 de *gag* hasta el gen *tat* y el fragmento 3 (F3) comprende desde *vpr* hasta la LTR en 3'.

Cebadores Amplificación Genoma Completo				
Fragmento	Cebador	Orientación	Localización (posición/región genómica)	Secuencia (5'-3')
1	187	+	57-77/LTR (U3)	CACACACAAGGCTACTTCCCT
1	147	-	2610-2635/pol (RT)	TCTTCTGTCAATGGCCATTGTTTAAC
Anidado 1	554	+	456-485/LTR (R)	GTCTCTCTTGTTAGACCAGATCTGAGCCTG
Anidado 1	352	-	2211-2238/gag (p6)	CAGTTCCTTGTCGCTCAGCTCCTGCTTC
2	148	+	1963-1985/gag (p7)	TTCAATTGTGGCAAAGAAGGGCAC
2	101	-	5891-5913/tat	TGGAAGCCTGCAGGAAGTCAGCC
Anidado 2	136	+	2168-2188/gag (p6)	AAGAGAGCTTCAGGTTTGGGG
Anidado 2	33	-	5825-5844/vpr-tat	TCTACTGGCTCCATTTCTTG
3	294	+	5677-5700/vpr	ATGGCTTAGGGCAACATATCTATG
3	163	-	9664-9687/LTR (U5)	CTGAGGGATCTCTAGTTACCAGAG
Anidado 3	551	+	5729-5760/vpr	GGAAGCCATAATAAGAATTCTGCAACAACCTGC
Anidado 3	555	-	9637-9669/LTR (U5)	CCAGAGTCACACAACCGCGGGGCACACTACT

Tabla 3.1- Cebadores utilizados en la amplificación de los genomas completos. Se indica el fragmento que amplifican, el nombre del cebador, su orientación (+ sentido y – antisentido), su localización en el genoma (numerada con respecto al virus HXB2(número de acceso #K03455)) y su secuencia.

Para realizar las diversas PCRs la enzima utilizada fue la Expand High Fidelity (ROCHE, 2,5 U de enzima por reacción), con su correspondiente tampón comercial a una concentración de Mg^{2+} de 1,5 mM. La concentración de desoxinucleótidos (dNTPs) (Amersham) empleada fue de 0,125 mM y la de cada uno de los cebadores utilizados fue de 2 ng/μl. Las reacciones se llevaron a cabo en un termociclador C-1000 (BioRad). Se utilizó esta enzima de alta fidelidad para minimizar los errores producidos durante la amplificación.

Para la amplificación se usaron los siguientes ciclos de PCR:

- | | | |
|------------------------------|---|------------|
| 1) 94°C 5 minutos. | } | 35 ciclos. |
| 2) 94°C 30 segundos. | | |
| 3) 55°C 30 segundos. | | |
| 4) 72°C entre 2 y 4 minutos. | | |
| 5) 72°C 10 minutos. | | |

3.4.1.3 Secuenciación del genoma completo.

Las secuencias del genoma completo de los virus en pase 1 y 11 fueron obtenidas en trabajos anteriores (Borderia y cols., 2010). Las secuencias globales del genoma completo después del pase 21 y 31 en gran tamaño poblacional se determinaron en las dos cadenas de ADN. Los productos de PCR obtenidos previamente (apartado anterior) fueron secuenciados empleando el equipo comercial “Big Dye™ Terminator Cycle Sequencing v2.0” (Applied Biosystems), siguiendo las instrucciones suministradas por el fabricante. Para la secuenciación de los fragmentos amplificados de cada clon viral, se usaron distintos cebadores tanto sentido como antisentido de los disponibles en el laboratorio.

La secuenciación se realizó mediante amplificación por PCR asimétrica en presencia de los cuatro dNTPs, marcados cada uno con un colorante fluorescente distinto.

La PCR asimétrica consistió en:

- 1) 94°C 3 minutos.
 - 2) 94°C 30 segundos.
 - 3) 55°C 15 segundos.
 - 4) 60°C 4 minutos.
- } 25 ciclos.

Las reacciones de secuenciación se realizaron en un termociclador C-1000 (BioRad). Los productos obtenidos fueron analizados en un secuenciador automático ABI PRISM™ modelo 3700 (Applied Biosystems) por la Unidad de Genómica del ISCH. Las secuencias generadas se validaron y analizaron utilizando el programa SeqMan (Lasergene), que permite comprobar el cromatograma obtenido por el secuenciador.

3.4.1.4 Obtención de las cuasiespecies virales.

El análisis del espectro de mutantes se realizó en 4 regiones genómicas, que se amplificaron con 3 pares diferentes de cebadores (Figura 3.3). Se seleccionaron estas cuatro regiones distintas para tener una mejor representación de la variabilidad del virus a lo largo de todo el genoma. El primer par de cebadores amplifica un fragmento comprendido entre las posiciones del genoma completo 672 a la 1363 (692 nucleótidos), numerados según el virus HXB2 (base de datos de Los Alamos), que comprendían la secuencia líder y la proteína p17 del gen *gag*. El segundo par amplifica un fragmento comprendiendo desde el nucleótido 6045 a la 6734 (690 nucleótidos), correspondientes a la proteína Vpu y la región V1-V2 del gen *env*. Este fragmento se dividió en las dos regiones mencionadas para obtener datos separados de ambos genes. El último par amplifica desde la posición 7039 a la 7641 (603 nucleótidos) que abarcaba la región V3-V4 del gen *env* (Tabla 3.2). En total, las 3 amplificaciones cubrían 1985 nucleótidos, que suponían alrededor del 20% del genoma completo del VIH-1.

Cebadores cuasiespecies			
Cebador	Secuencia (5'-3')	Orientación	Localización (posición/región genómica)
170	TCTCTAGCAGTGGCGCCCGAACAGGGAC	+	672-699/LTR
174	TCTATCCCATTCTGCAGCTTCCTCATTGAT	-	1334-1363/ <i>gag</i>
560	CTATGGCAGGAAGAAGCCGAGA	+	6045-6066/ <i>vpu</i>
52	TAATGTATGGGAATTGGCTCAA	-	6713-6734/ <i>env</i>
27	ATAAGCTTGCAGTCTAGCAGAAGAAGA	+	7039-7065/ <i>env</i>
167	TTCTCCAATTGTCCCTCATATCTCCTCTCCA	-	7610-7641/ <i>env</i>

Tabla 3.2 Cebadores empleados en la obtención de las cuasiespecies virales. Se indica el nombre del cebador, su secuencia, orientación (+ sentido y – antisentido) y localización en el genoma (numerada con respecto al virus HXB2).

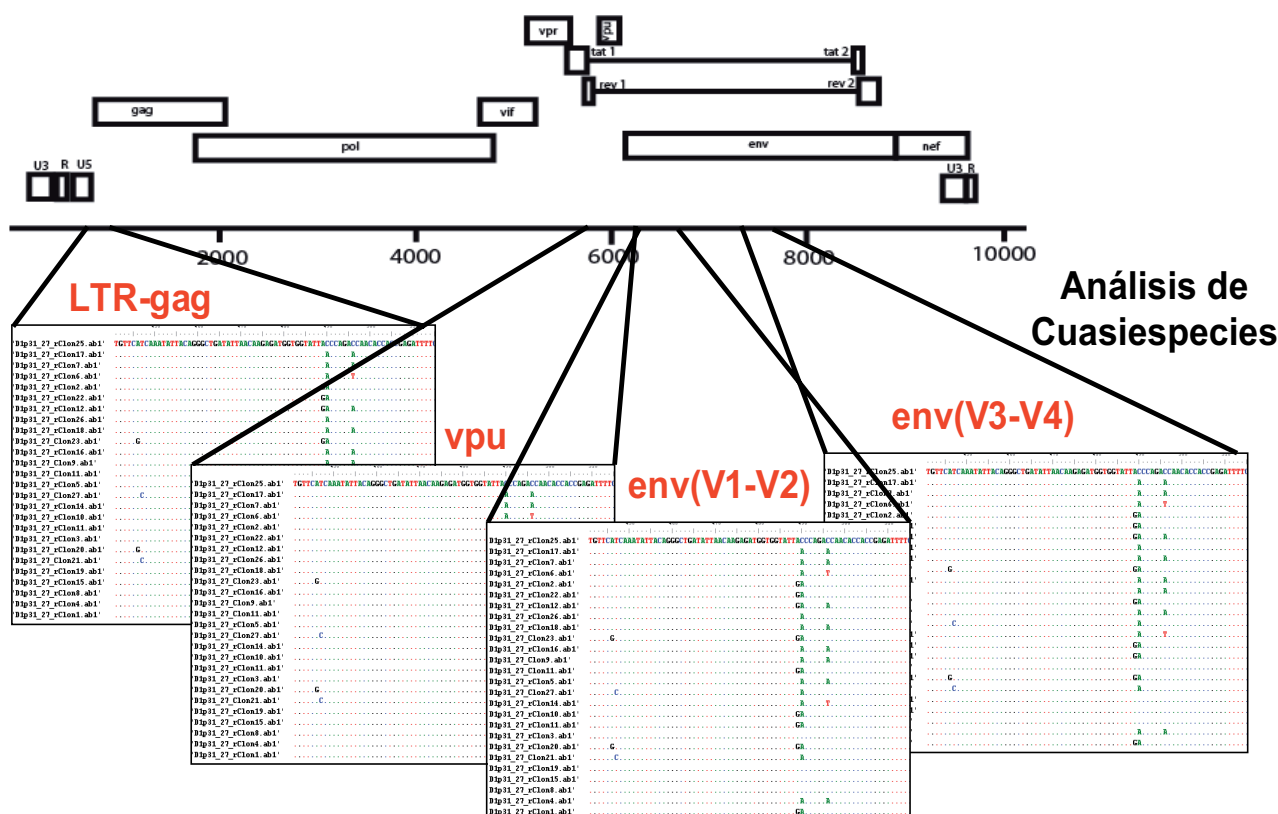


Figura 3.3 Esquema de las regiones genómicas en las que se realizó el análisis de las cuasiespecies virales. Se indican las 4 regiones en las que se obtuvo el espectro de mutantes.

Posteriormente, los productos de PCR se clonaron utilizando el sistema de clonaje TA TOPO cloning (Invitrogen), siguiendo las instrucciones del fabricante. Una vez realizado el clonaje se analizaron alrededor de 20 clones independientes para representar las cuasiespecies de las poblaciones virales en cada pase.

Finalmente, las secuencias de estos clones fueron utilizadas para la caracterización y cuantificación de la heterogeneidad genética del espectro de mutantes. La secuenciación se realizó de igual manera que para el genoma completo, empleando uno de los cebadores utilizados en la amplificación. En el caso de las cuasiespecies con los cebadores 170/174 en la región de la secuencia leader/gag se empleó para secuenciar el cebador 170; para las cuasiespecies amplificadas con los cebadores 27/167 en la región *vpu/env* (V1-V2) el utilizado fue el cebador 167 y, finalmente, para las cuasiespecies con los cebadores 560/52 en la región *env* (V3-V4), el cebador 560 (ver Tabla 3.2).

3.4.2 Anál isis de las secuencias nucleotídicas.

3.4.2.1 Alineamiento de los genomas completos.

Para el análisis de mutaciones y de la filogenia de los virus durante los pases seriados en gran tamaño poblacional, se alinearon las secuencias de los genomas completos de cada virus en los diferentes pases, con respecto a las de pases anteriores y con las de la totalidad de los virus estudiados. Para ello se utilizaron programas de alineamiento múltiple como ClustalX 2.0 (Larkin y cols., 2007) y MUSCLE (Edgar, 2004). Las secuencias de cada una de las cuasiespecies obtenidas se alinearon de igual forma.

3.4.2.2 Detección de las mutaciones y deleciones aparecidas.

Utilizando los alineamientos de las secuencias de cada uno de los virus en los diferentes pases, se analizaron las mutaciones aparecidas durante los pases en gran tamaño poblacional. Con este objetivo, se compararon los virus en cada pase con su correspondiente virus inicial. Para la localización de las mutaciones y la comprobación de su asociación a un cambio o no de aminoácido, se dividieron los genomas completos en los alineamientos de los diferentes genes y se tradujeron a proteínas. En este proceso se empleó la aplicación GeneCutter (http://www.hiv.lanl.gov/content/sequence/GENE_CUTTER/cutter.html) disponible en las base de datos de Los Alamos. Así, se pudieron numerar las diferentes mutaciones dentro de cada gen, teniendo como referencia el virus HXB2 y se detectaron los cambios de aminoácido asociados.

Además, se examinó la aparición de deleciones en la región V1-V2 del gen *env*, que previamente se habían detectado en las secuencias globales. Para ello, se volvió a utilizar la técnica de GeneScan, con la que se podía calcular la frecuencia de cada una de los variantes delecionados presentes en la población. En este caso, solamente se utilizó el ADN proveniente de estos virus, amplificándose de nuevo con los mismos cebadores utilizados en los ensayos de competición, ya que abarcaban la zona de interés. Se siguió el mismo procedimiento que en las competencias y se calcularon las frecuencias de las diferentes deleciones dentro de la población midiendo el área de los picos obtenidos de cada uno de los variantes presentes en la población.

3.4.2.3 Anál isis de la presión selectiva operando en los diferentes genes en el genoma completo.

Para contrastar el tipo de fuerzas selectivas que estaban operando a lo largo todo el genoma durante los pases de recuperación, se realizó un análisis por Máxima Verosimilitud de la selección natural codón a codón en cada uno de los alineamientos con todos los virus en los distintos genes. Se utilizó el programa HyPhy (Pond, Frost y Muse, 2005) para estimar el número de sustituciones sinónimas (S) y no sinónimas (N) en cada sitio asumiendo el modelo de sustitución de Hasegawa-Kishino-Yano (HKY). A partir de estas estimaciones, se calculó el estadístico dN-dS, donde dN era el número de N por posibles sitios no sinónimos y dS el número de S por posibles sitios sinónimos.

El mencionado estadístico fue utilizado para detectar los codones bajo selección positiva, en los que el valor de dN-dS era significativamente positivo al nivel del 5%, y bajo selección negativa en los que ocurría lo contrario. Finalmente, se calculó el porcentaje de codones que se apartaban significativamente de la neutralidad en los diferentes genes. Conociendo el número de codones bajo selección positiva y negativa, se examinó la preponderancia de una u otra fuerza en cada uno de los genes para establecer el tipo de selección predominante a lo largo del proceso de recuperación.

3.4.2.4 Estimación del modelo evolutivo seguido por las secuencias.

Las secuencias obtenidas también se utilizaron para estimar el modelo evolutivo seguido por los virus estudiados. Esta estimación se realizó por el método de Máxima Verosimilitud, empleando para ello el programa jModelTest (Posada, 2008). Con esta evaluación se obtenía el modelo o modelos que mejor se ajustaban a la evolución de las secuencias. Para seleccionar el mejor modelo, se empleó el criterio de información de Akaike (AIC) utilizando los valores de verosimilitud asociados a cada uno de los modelos. Además, se obtuvieron los diferentes parámetros asociados al mejor modelo y los parámetros medios correspondientes a la media ponderada de todos los modelos probados.

En primer lugar, se utilizó el alineamiento que contenía todos los genomas globales obtenidos de los virus estudiados para estimar el modelo seguido por los virus a lo largo de los pases de recuperación de eficacia. Posteriormente, se estimaron los modelos seguidos por las cuasiespecies en cada una de las regiones analizadas, con el objetivo de aplicar los modelos obtenidos en los análisis subsiguientes de heterogeneidad de las cuasiespecies.

3.4.2.5 Cálculo de la heterogeneidad de las cuasiespecies virales.

Con los parámetros evolutivos medios estimados en cada una de las regiones, se procedió a calcular las distancias genéticas dentro de cada cuasiespecie. Con este objetivo, se utilizó el método de Máxima Verosimilitud empleando los parámetros medios de los modelos obtenidos por el jModeltest. Al utilizar esta media ponderada por el peso de cada modelo, en lugar de únicamente los del mejor modelo, se estaban utilizando unos parámetros mejor ajustados a los datos. Este cálculo se realizó con el programa PAUP (Swofford, 1998). La heterogeneidad de las cuasiespecies fue medida como el valor medio de la distancia genética, medida en número de sustituciones por sitio, comparando dos a dos todos los clones de la cuasiespecie. Como en todas las regiones analizadas el mejor modelo era el mismo, el “General Time Reversible” con distribución Gamma de mutaciones (GTR+G), y no se encontraron diferencias significativas entre las distancias genéticas dentro del mismo virus en cada región (análisis de varianzas de una sola vía (ANOVA) con medidas repetidas p -valor >0.05), se decidió utilizar los valores medios de distancia genética de las cuatro regiones de cada virus estudiado para los análisis posteriores.

3.4.2.6 Análisis de mutaciones sinónimas y no sinónimas dentro de las cuasiespecies.

Con los alineamientos de las cuasiespecies virales se calculó el número de sustituciones

sinónimas y no sinónimas presentes en el espectro de mutantes comparándolas con el pase inicial de cada virus. Este cálculo se llevó a cabo utilizando el método modificado de Nei-Gojobori (Nei y Gojobori, 1986) implementado en el programa MEGA 5.0 (Tamura y cols.). Los valores calculados fueron los de dS y dN, para los cuales además se asumió el modelo de Jukes-Cantor. Los valores medios de las cuatro regiones de cada virus de dN y dS fueron utilizadas para examinar las diferencias estadísticas.

3.5 Análisis filogenéticos de las secuencias virales obtenidas.

3.5.1 Estimación filogenética por Máxima Verosimilitud.

Utilizando el método de Máxima Verosimilitud, se estimó la filogenia de las secuencias globales correspondientes a los genomas completos. Para ello, de nuevo se utilizaron los parámetros medios de los modelos estimados por jModelTest y se procedió a la estimación del árbol filogenético que mejor describía las relaciones filogenéticas de las secuencias. Como grupo externo para enraizar el árbol se utilizó el virus s61 original del que descendían todos los virus estudiados. Como medida del soporte estadístico de cada rama se utilizó el método de razón de verosimilitud aproximado (aLRT) (Anisimova y Gascuel, 2006). Toda esta estimación se llevó a cabo empleando el programa PHYML (Guindon y Gascuel, 2003). Una vez obtenido el árbol, se calculó la distancia genética de cada virus a la raíz del mismo, utilizando para ello el programa TreeStat (<http://tree.bio.ed.ac.uk/software/treestat/>).

3.5.2 Análisis de las tasas de evolución de los distintos virus mediante estimación bayesiana.

Para la estimación de las tasas de evolución de cada uno de los virus se emplearon métodos bayesianos. Se utilizó el programa Beast (Drummond y Rambaut, 2007) y sus programas accesorios para llevar a cabo las mencionadas estimaciones. En dicho programa se tuvieron en cuenta los pases, fijándose como unidad de tiempo cada pase realizado. Tras realizar varios análisis, en los que se probaron diferentes parámetros *a priori*, se fijó como modelo evolutivo el GTR con sitios invariantes y distribución gamma de mutaciones (GTR+I+G), se adoptó un reloj molecular relajado con una distribución logNormal, que era el que mejor funcionaba, para calcular las diferentes tasas de evolución y se asumió una población constante. Se iteraron dos cadenas de Markov de 20 millones de generaciones, llegando ambas a valores de tamaño de población efectiva mayores de 100 en todos los parámetros, combinándose después para obtener un fichero con 36.000 árboles en total, ya que se eliminaron los primeros 2.000 de cada cadena para generar el árbol consenso con aquellos árboles obtenidos en la fase estable de la búsqueda. El archivo combinado de los árboles estimados se utilizó a su vez para calcular las tasas de evolución de las ramas externas, que correspondían a los virus estudiados. Este cómputo se llevó a cabo con el programa TreeStat.

3.5.3 Redes filogenéticas de las cuasiespecies virales.

Para el estudio de la dinámica poblacional dentro de las cuasiespecies, se empleó el análisis por redes de los alineamientos de los clones de cada virus obtenidos en las diferentes regiones del genoma. Estos análisis se realizaron utilizando diferentes métodos para generar redes de distancias genéticas, que se diferenciaban en el modo de asignar las distancias entre las secuencias. Los métodos elegidos fueron los siguientes: Median Network, basado en caracteres, Parsimony Splits, basado en parsimonia y NeighbourNet, basado en distancias. Todos estos métodos están disponibles en el programa SplitsTree4 (Huson y Bryant, 2006).

3.6 Caracterización fenotípica de los virus obtenidos en los cultivos celulares.

3.6.1 Titulación de los virus.

El título de los virus analizados se obtuvo calculando el número de dosis infectivas 50 por ml (TCID₅₀/ml). Para ello, se infectaron en placas de cultivo de 96 pocillos, 1 x 10⁵ células MT-2 por pocillo, con diluciones seriadas de cada clon viral (desde 10⁻¹ hasta 10⁻⁸) en 100 µl de medio completo. De cada una de las diluciones se realizaron 8 réplicas. A los 7 días se observó la aparición de efecto citopático (ECP), que consiste en la formación de sincitios, es decir, células gigantes multinucleadas. Se contó el número de pocillos positivos para ECP por dilución y réplica en la que aparecían, y se obtuvo el título viral según la fórmula de Spearman-Kärber (Hubert, 1984). De cada virus estudiado se realizaron 2 réplicas de la titulación.

3.6.2 Determinación de los niveles de proteína p24 viral.

Los niveles de proteína p24 se cuantificaron en los sobrenadantes de las infecciones en cada pase o de los distintos mutantes. Se utilizó el método enzimático automatizado de medida de antígeno p24 (HIV Ag, ROCHE) con el analizador Elecsys 2010.

3.6.3 Ensayo de la citopatogenicidad de los virus estudiados.

Para conocer la citopatogenicidad de los virus, se analizaron cultivos celulares infectados. Con este propósito, 1 x 10⁵ células MT-4 en 100 µl de medio completo fueron infectadas con cada virus a una m.o.i. de 0,01. Todas las infecciones fueron iniciadas a la vez en placas de 96 pocillos (p96) con el objeto de evitar diferencias debidas al estado de las células. Con cada virus se realizaron infecciones por duplicado. Durante la infección de los cultivos, se realizó un seguimiento de los mismos utilizando microscopía óptica para observar la aparición de sincitios, disgregación celular, apoptosis u otros signos del avance de la infección. Los valores de viabilidad de cada cultivo se calcularon por exclusión de azul de Tripan utilizando cámaras de Neubauer. Cuando los cultivos

presentaban una viabilidad por debajo del 40% se recogían. Los valores de viabilidad observados en el día 3 de cultivo fueron los utilizados para probar las diferencias entre los distintos virus y realizar análisis estadísticos.

3.6.4 Actividad Retrotranscriptasa.

El ensayo de actividad retrotranscriptasa (RT) se llevó a cabo utilizando una reacción de extensión de poli-Adenina “*in vitro*” usando marcaje radiactivo con fósforo-32 (^{32}P). Se empleó un cóctel de RT con Tris pH 7,8, KCl 3M, DTT 0,1M, MgCl_2 0,15M, Poly A 1mg/ml, dT 1/20 y NP-40 2%. A esta mezcla se le añadió 1 μl /ml de ^{32}P y se utilizaron 50 μl del cóctel con isótopo por pocillo en placas p96. Posteriormente, se adicionaron 10 μl de sobrenadante de infecciones virales y se incubó durante 1 hora a 37°. Después de la incubación, se añadieron 6 μl de cada pocillo a un papel de 3mm y se dejó secar. Se realizaron 4 lavados de 5 minutos con citrato de sodio salino (SSC) 2X y dos lavados de 1 minuto con etanol absoluto, cuantificándose la señal radiactiva mediante un lector Phosphorimager (Fuji BAS).

3.7 Análisis por regresión múltiple.

Todas las variables analizadas en los distintos virus fueron utilizadas para generar un modelo de regresión lineal múltiple de la recuperación de eficacia del VIH-1 “*in vitro*”. Inicialmente todas las variables estudiadas se añadieron al modelo, probándose el modelo incluyendo o no una constante. La variable dependiente, como es lógico, fue la eficacia biológica. Se ajustó un modelo lineal y se realizaron pruebas ANOVA para comprobar la significancia del modelo y de correlación múltiple para comprobar cuáles de las variables tenían parámetros asociados significativamente con la eficacia biológica. Además, se realizaron pruebas de colinearidad para evitar incluir variables linealmente dependientes. Se estimaron R^2 y R^2 ajustadas del modelo para examinar su comportamiento en la explicación de la variabilidad observada en la eficacia, así como cálculos del error estándar de las estimaciones. Finalmente, para hacer el modelo lo más sencillo posible, se fueron eliminando aquellas variables que no estaban asociadas significativamente a la eficacia en la regresión múltiple y que al eliminarlas no hacían empeorar el modelo, utilizando el método “hacia atrás” o backward. De esta forma se obtuvo el modelo mejor y más simple posible, lográndose distinguir las variables más asociadas con la recuperación de eficacia. Todo este análisis fue llevado a cabo utilizando el programa PasW Statistics 18 y el paquete R.

3.8 Análisis del papel de cambios de aminoácido aparecidos en el proceso de pérdida y recuperación de eficacia biológica mediante mutagénesis dirigida.

3.8.1 Obtención de secuencias de ADN con mutaciones aparecidas en el proceso de pérdida y recuperación de eficacia biológica.

Para obtener ADN viral con algunas de las mutaciones más interesantes aparecidas en los pases de cuello de botella y en gran tamaño poblacional, se utilizaron como molde dos plásmidos que contenían cada uno la mitad del genoma proviral clonado del virus original s61. El plásmido pFAM4 contenía la región 5' del genoma, desde la 5'LTR hasta un sitio EcoRI en el gen *vpr*, y el plásmido p61FC comprendía la región 3', desde el mismo sitio EcoRI hasta la 3'LTR (Olivares, Shaw y Lopez-Galindez, 1997) (Figura 3.4). Utilizando estos dos plásmidos como molde, se llevaron a cabo diferentes PCRs con cebadores modificados para introducir la mutación deseada. Se utilizó uno u otro plásmido según la localización de la mutación y los correspondientes cebadores para cada cambio (Tabla 3.3).

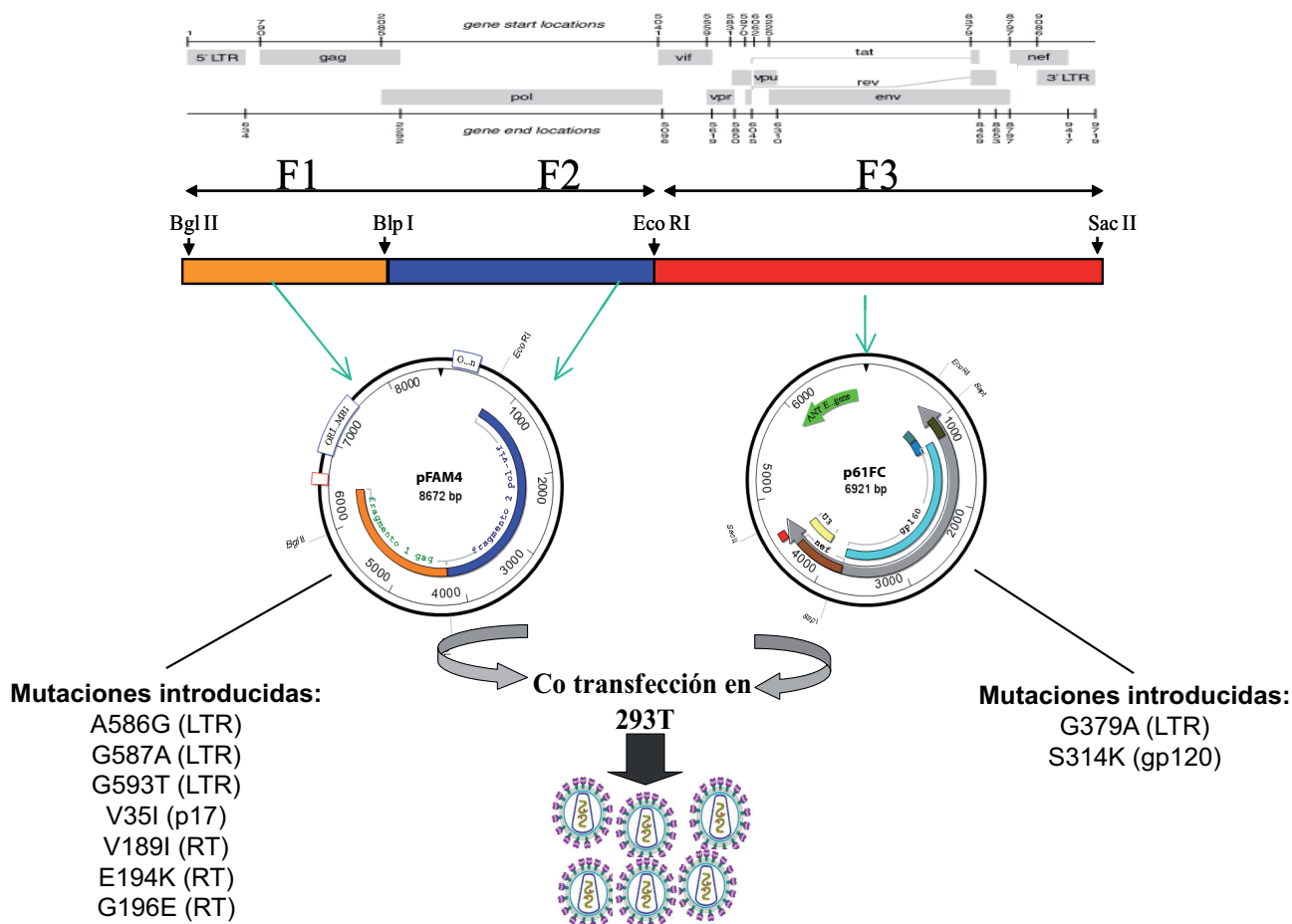


Figura 3.4 Esquema del clonaje llevado a cabo para realizar la mutagénesis dirigida. El genoma completo del virus s61 se dividió en dos fragmentos, cada uno contenido en un plásmido diferente. Cada uno de estos dos fragmentos se utilizó como molde para introducir las mutaciones deseadas. Se indica el fragmento que se utilizó para cada una de las mutaciones estudiadas según su localización. Una vez obtenidos los dos fragmentos insertados en los plásmidos correspondientes, éstos se co-transfectaron en células 293T para obtener los virus mutantes. Figura modificada de una cedida por la Dra. M. Pernas.

Cebadores utilizados en mutagénesis dirigida		
Gen/Cambio de aminoácido	Cambio de nucleótido (numerado según s61)/Dirección	Secuencia del cebador
LTR/NC	G379A/Sentido	5'-CCGCTGGGGACTTTCCAGGGAAGCGTGGCCTGGGCGGGACTGGGG-3'
LTR/NC	G379A/Antisentido	5'-CCCCAGTCCCGCCCAGGCCACGCTTCCCTGGAAAGTCCCCAGCGG-3'
LTR/NC	A586G/Sentido	5'-GCCCCGTCTGTTGTGTGACTCTGATGGCTAGAGATCCCTCAGACC-3'
LTR/NC	A586G/Antisentido	5'-GGTCTGAGGGATCTCTAGCATCAGAGTCACACAACAGACGGGC-3'
LTR/NC	G587A/Sentido	5'-GCCCCGTCTGTTGTGTGACTCTGATAACTAGAGATCCCTCAGACC-3'
LTR/NC	G587A/Antisentido	5'-GGTCTGAGGGATCTCTAGTTATCAGAGTCACACAACAGACGGGC-3'
LTR/NC	G593T/Sentido	5'-GCCCCGTCTGTTGTGTGACTCTGATAGCTAGATATCCCTCAGACC-3'
LTR/NC	G593T/Antisentido	5'-GGTCTGAGGGATATCTAGCTATCAAGAGTCACACAACAGACGGGC-3'
gag (p17)/V35I	G892A/Sentido	5'-ATAGATTAACAAACATTTAATATGGGCAAGCAGGGAGCTAGAACG-3'
gag (p17)/V35I	G892A/Antisentido	5'-CGTTCTAGCTCCCTGCTTGCCCATATAAATGTTTAAATCTAT-3'
pol (RT)/V189I	G3123A/Sentido	5'-CATGGACGATTTGTATATAGGATCTGACTTAGAAATAGGGCAGC-3'
pol (RT)/V189I	G3123A/Antisentido	5'-GCTGCCCTATTTCTAAGTCAGATCCTATATACAAATCGTCCATG-3'
pol (RT)/E194K	G3138A/Sentido	5'-CGATTTGTATGTAGGATCTGACTTAAATAGGGCAGCATAGA-3'
pol (RT)/E194K	G3138A/Antisentido	5'-TCTATGCTGCCCTATTTTAAGTCAGATCCTACATACAAATCG-3'
pol (RT)/G196E	G3145A/Sentido	5'-CGATTTGTATGTAGGATCTGACTTAGAAATAGAGCAGCATAGA-3'
pol (RT)/G196E	G3145A/Antisentido	5'-TCTATGCTGCTCTATTTCTAAGTCAGATCCTACATACAAATCG-3'
env (gp120)/S314K	G7170A-T7171A/Sentido	5'-ACAAGGAGAAAATACATGTAGGACATGTAGGACCAGGCAGAGC-3'
env (gp120)/S314K	G7170A-T7171A/Antisentido	5'-GCTCTGCCTGGTCTACATGTCCTACATGTATTTTCTCCTTGT-3'

Tabla 3.3 Cebadores utilizados para la mutagénesis dirigida. Se indica la localización en el genoma de cada cebador, el cambio aminoacídico que suponen (NC cuando la mutación es en zona no codificante), el cambio nucleotídico, numerado con respecto al virus s61, su dirección y la secuencia de cada uno. Se muestran en rojo las mutaciones introducidas.

La reacción de PCR se llevó a cabo utilizando 50 ng/μl de molde y la enzima de alta fidelidad de copia Pfu Turbo™ (Stratagene 2.5 U/μl) en un termociclador C-1000 (BioRad) con los siguientes ciclos de reacción:

1. 95°C 3 minutos.
 2. 95°C 30 segundos.
 3. 50°C 1 minuto.
 4. 68°C 18 minutos.
 5. 15°C 10 minutos.
- } 15 ciclos.

Una vez obtenidas las PCRs, se digirió el ADN con la enzima DpnI (Stratagene, 10 U/μl) durante 1 hora a 37°C con el fin de eliminar el ADN metilado. Al final de este proceso ya se disponía de ADN viral con las mutaciones que se querían estudiar incorporadas.

Se transformaron 100 μl de bacterias E.Coli (DH5α) competentes con 2 μl del ADN digerido y se incubaron durante 30 minutos a 4°C. Posteriormente, se aplicó un choque térmico de 1 minuto y 30 segundos a 42°C, seguido de 2 minutos a 4°C. Se añadieron 800 μl de medio LB con Ampicilina (75 μg/ml) y se incubaron durante 1 hora a 37°C en agitación. Finalmente, se plaquearon 100 μl del cultivo en placas p100 con agar y LB-Ampicilina, manteniéndose toda la noche a 37°C.

De las colonias obtenidas, se seleccionaron 5 para amplificar el ADN correspondiente a los plásmidos introducidos. Cada uno de estos plásmidos se secuenció para comprobar la presencia de la mutación deseada y se seleccionó una de cada mutante para crecer en Maxiprep utilizando el kit Quia Plasmid Maxi (QUIAGEN), siguiendo las instrucciones del fabricante. La cuantificación del ADN obtenido se realizó en un aparato NanoDrop (Termo Scientific).

3.8.2 Generación y obtención de los virus mutados.

El primer paso para la generación de los virus mutados fue la digestión, con la enzima EcoRI (New England BioLabs, 20 Unidades), de 20 µg del plásmido pFAM4 y de la misma cantidad de p61FC, portadores uno u otro en cada caso de los cambios a estudiar (Figura 3.4), con el fin de favorecer la recombinación homóloga tras la transfección. Tras las digestiones, los plásmidos fueron purificados utilizando fenol:cloroformo y precipitados con NaAc (0,3 M) y etanol al 70%.

Posteriormente, se procedió a la transfección de 3×10^6 células 293T en 6 ml de medio DMEM completo. Se añadió el ADN con CaCl_2 2 M y HEPES salino tamponado (HBS, en inglés) (pH 7,05) utilizando la técnica de burbujeo y se incubaron las células 24 horas a 37°C. A las 24 horas se cambió el medio y a las 72 horas se añadieron 5×10^6 células MT-4 para amplificar la semilla viral inicialmente generada. Cuando la viabilidad de los cultivos cayó por debajo del 40% se recogieron los cultivos, obteniéndose de esta forma sobrenadantes con los virus mutados en las posiciones a estudiar. Finalmente, con dichos virus se realizaron ensayos de competición, citopatogenicidad, producción de p24 y titulación para llevar a cabo su caracterización.

3.8.3 Modelización de las estructuras mutadas.

Al disponer de las secuencias correspondientes a cada proteína mutada mediante mutagénesis dirigida, se procedió a modelar su estructura para comprobar los posibles cambios que producían las mutaciones introducidas. De igual modo, se intentaron modelar las estructuras de ARN correspondientes a la región LTR de los mutantes en dicha región.

Para el modelado de la estructura de ARN de las LTRs, se utilizó el programa de predicción de estructura secundaria de ARN GeneBee (Brodsky y cols., 1992). Para confirmar las estructuras obtenidas se utilizó el programa Vienna RNA Package (Hofacker, 2003).

También se modeló la estructura de las diferentes proteínas modificadas. En este caso, se utilizó en primer lugar el programa Blast (Altschul y cols., 1990) para encontrar la estructura más parecida a cada una de las proteínas a estudiar presente en la base de datos de estructuras PDB. Esta estructura se utilizó como molde para modelizar las proteínas mutadas, mediante el programa DeepView (Guex y Peitsch, 1997) y el servidor en red SwissModel (<http://spdbv.vital-it.ch/>). Finalmente, se estudiaron dichas estructuras utilizando los programas RasMol (Sayle y Milner-White, 1995) y PyMol (<http://www.pymol.org>).

3.9 Generación de paisajes adaptativos.

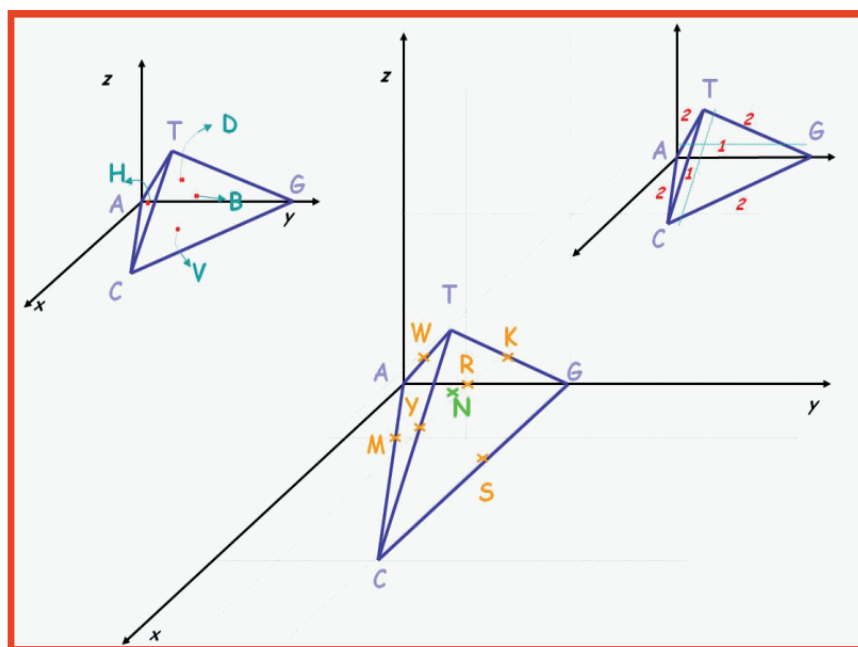
3.9.1 Obtención de un paisaje adaptativo por técnicas filogenéticas.

La primera aproximación para generar un paisaje real de eficacias de variantes del VIH-1 obtenidas “in vitro” se llevó a cabo utilizando técnicas filogenéticas convencionales. En este análisis se incluyeron tanto las secuencias de los virus iniciales como las de los virus después de los pases seriados en cuello de botella y a lo largo de los pases de recuperación (Figura 3.1). Además, se añadió el virus original s61 del que todos descendían. Tras alinear dichas secuencias, se estimó el árbol filogenético que mejor describía las relaciones evolutivas de dichas secuencias, utilizando para ello el método de Máxima Verosimilitud. La estimación del árbol se llevó a cabo utilizando el programa PHYLML con los parámetros medios de los modelos obtenidos por jModelTest, siguiendo la misma metodología que en el apartado 3.5. El árbol obtenido permitió derivar coordenadas bidimensionales utilizando las distancias genéticas estimadas. Cada punto XY correspondiente a una secuencia fue etiquetado con su correspondiente eficacia biológica asociada, con lo que se obtenía una representación con coordenadas XYZ. Posteriormente, se hizo uso del programa de generación de imágenes matriciales SurGe (<http://surgeweb.sweb.cz/>) para generar una matriz de puntos derivados de los datos y obtener una representación en 3D de un paisaje de eficacias filogenético. Para comprobar la robustez de la técnica, el mismo procedimiento se llevó a cabo utilizando diferentes métodos filogenéticos para obtener el árbol inicial. Así, las distancias genéticas también se obtuvieron por Neighbour Joinning y Median Network, generando paisajes de topología muy similar a la del de Máxima Verosimilitud.

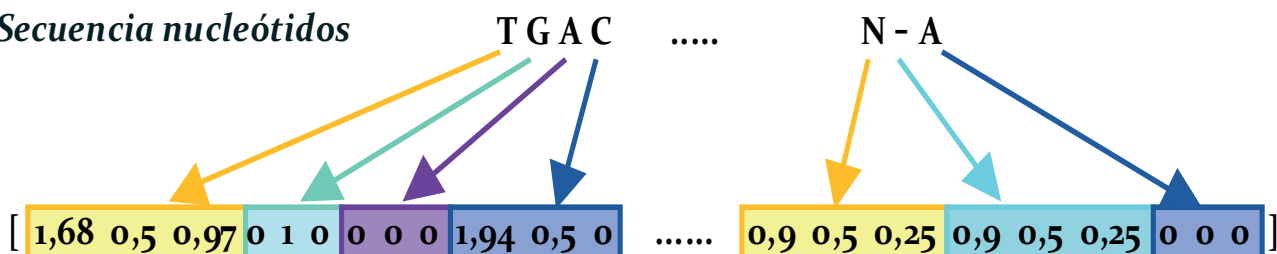
3.9.2 Obtención de un paisaje adaptativo utilizando redes neuronales.

En la clasificación de datos mediante ANN, se necesita entrenar previamente las redes utilizadas. En dicho entrenamiento de los mapas SOM, en primer lugar se transformaron las cadenas de nucleótidos en vectores numéricos. Para ello, se necesitaba codificar 15 letras distintas, que se correspondían con los nucleótidos Adenina (A), Guanina (G), Citosina (C), Timina (T) y las letras asociadas a las posibles combinaciones de estas: M (A o C), R (A o G), S (C o G), W (A o T), Y (C o T), K (T o G), H (A, C o T), V (A, C o G), D (G, A o T), B (G, C o T) y N (A, G, C o T). Además, se podía encontrar el carácter (–) que representa una delección en determinadas posiciones de la cadena de ADN. A priori, se buscó una codificación que permitiese que la distancia Euclídea entre cada par de letras fuese constante para todas ellas. Esto llevó a utilizar el concepto de tetraedro, asociando cada letra AGCT con los 3 números correspondientes a las coordenadas de los vértices de un tetraedro tridimensional. El resto de letras se codificaron con las coordenadas de los puntos medios de las aristas o los planos (Figura 3.5). La letra N, así como la delección (no se le dio un tratamiento especial), se codificaron con las coordenadas del centro del tetraedro. Se optó por una codificación en tetraedro 3D irregular (Delgado, 2009). En esta codificación, se consideraba que existen mutaciones más probables que otras. En particular, las transiciones A-G y C-T son más probables que las transiciones A-C, A-T, C-G y G-T, por lo que se podía usar un tetraedro que situase A, G y T en la base del tetraedro

y C en la cima, estando a distancia 1 los vértices A y T de G y C, respectivamente, y a distancia 2 todas las demás combinaciones (Figura 3.5). Además, se comprobó que, en general, la existencia de una determinada letra en la misma posición de todos los patrones de entrenamiento no influía en el resultado final de organización de los vectores sinápticos de la red. Por esta razón, en los experimentos que se realizaron se optó por la opción de trabajar con las cadenas sin repetición. Esto permitía reducir la dimensión de cada patrón de caracteres.



Secuencia nucleótidos



Vector numérico

Figura 3.5 Codificación utilizada para convertir las secuencias nucleotídicas en vectores numéricos. Se utilizó una codificación basada en un tetraedro irregular, en el que los diferentes cambios entre nucleótidos no tienen la misma probabilidad. La N (combinación de cualquier nucleótido y la delección se situaron en el centro del tetraedro, mientras que las otras combinaciones de nucleótidos se colocaron en la mitad de las aristas que unen a cada nucleótido. Esta codificación fue desarrollada en Delgado y cols. Figura cedida por la Prof. S. Delgado.

Los parámetros de entrenamiento que se emplearon en las redes fueron:

- Arquitectura: 15 (filas) x 15 (columnas) de neuronas en la capa de salida
- Condición de parada: 2700 iteraciones (50 epochs)
- Coeficientes de aprendizaje: $a_1(1-t/a_2)$. $a_1 \rightarrow 0,1$; $a_2 \rightarrow 2700$; siendo t el tiempo
- Vecindad inicial: 14 x 14. Reducción de vecindad cada 54 iteraciones (1 epoch)

Posteriormente, se llevó a cabo un proceso de etiquetado en el que a cada neurona se le asociaba un valor de eficacia. Para ello, cada red de Kohonen entrenada fue sometida a un proceso de etiquetado supervisado, que se puede denominar “por ámbito”. Dada una red auto-organizativa y el conjunto de patrones de entrenamiento etiquetados con el valor eficacia que tenían asociado, cada neurona de la capa de salida fue etiquetada con un valor numérico obtenido a partir de la siguiente ecuación:

$$valor = \frac{\sum_{i=1}^L v_i f_i}{N}, \quad (1)$$

donde:

$$N = \sum_{i=1}^L v_i, \quad (2)$$

$$v_i = \frac{1}{\|w - x_i\|}, \quad (3)$$

f_i es el valor de la eficacia asociado al patrón de entrenamiento i , w es el vector sináptico de la neurona y x_i es el patrón i . Dado un valor concreto L (ámbito), se obtienen los L patrones más cercanos a cada neurona (en distancia Euclídea con el vector sináptico de la misma). Usando los valores de eficacia de estos L patrones, el vector de los L patrones y el vector sináptico de la neurona, se obtuvo un valor concreto con el que se etiquetó cada neurona de la red. A partir de estas redes etiquetadas por ámbito se generaron los gráficos que nos mostraban la superficie de eficacia.

Una vez se disponía de esta red, se podía utilizar para estimar la eficacia de secuencias no empleadas en el entrenamiento de la red. Al procesar la secuencia por la red se activaba la neurona cuyo vector sináptico se pareciese más (midiendo esta similitud en distancia Euclídea). El valor de eficacia biológica que se le asignaba era el asociado la neurona ganadora.

Para validar la calidad de estimación de eficacia obtenida por la red, se cuantificó el error de estimación que producía. Para ello, se entrenaron las redes de Kohonen con una secuencia menos de las utilizadas para generar el paisaje. Una vez entrenada cada red, se procedió a su etiquetado por eficacia y se procesó el patrón no utilizado en el entrenamiento, obteniéndose así una estimación de eficacia para el mismo. Este procedimiento se llevó a cabo reiteradamente, quitando cada vez una secuencia, hasta haber estimado todas. Así, se calculó el error medio que se cometía en la estimación de eficacia de secuencias no utilizadas para generar el paisaje.

3.10 Análisis estadísticos.

Para los diferentes análisis estadísticos empleados a lo largo de este estudio se utilizaron los programas estadísticos GraphPad, PasW Statistics 18 y el paquete R.

Las comparaciones entre los valores por grupos se realizaron mediante ANOVA con medidas repetidas. Una vez realizadas estas pruebas, las comparaciones dos a dos se comprobaron

con un *p-valor* corregido por el método de Bonferroni, teniendo en cuenta la varianza total.

Por otro lado, las correlaciones lineales simples se realizaron con métodos paramétricos (Pearson) y no paramétricos (Spearman) con el objeto de analizar relaciones entre variables dos a dos.

Además, en el caso de la correlación entre la eficacia y la heterogeneidad se utilizaron también modelos lineales de efectos mixtos para tener en cuenta los pases. Se tomaron los pases como medidas repetidas y se utilizó el método de Máxima Verosimilitud para estimar los modelos. Se probaron diferentes modelos con diferentes estructuras en la matriz de correlación y teniendo o no en cuenta los efectos aleatorios entre sujetos. Finalmente, se eligió el modelo con mayor verosimilitud utilizando AIC, en el que se estimó la significancia de la correlación entre las variables.

4. Resultados

4.1 Análisis de la recuperación de la eficacia biológica de clones biológicos del VIH-1.

4.1.1 Crecimiento continuo de la eficacia biológica global en clones biológicos del VIH-1 durante 30 pases seriados en gran tamaño poblacional.

Estudios previos llevados a cabo en el Servicio de Virología Molecular dieron como resultado un incremento de la eficacia biológica de virus previamente debilitados por efecto del Trinquete de Muller en tan sólo 10 pases en gran tamaño poblacional (Yuste y cols., 2005). Para estudiar el proceso de recuperación cuando estos pases se ampliaban, se realizaron 20 pases más con las mismas poblaciones virales y se analizaron exhaustivamente las propiedades biológicas de dichas poblaciones.

En primer lugar, se llevaron a cabo competiciones de todos los virus de los pases 21 y 31 frente al virus de referencia del laboratorio J1. Este virus es utilizado debido a que presenta una delección de 45 nucleótidos en la región V1-V2 del gen *env*, que le permite ser diferenciado de los demás virus en ensayos de GeneScan. El primer resultado obtenido tras estos experimentos fue una subida continuada de la eficacia biológica global a lo largo de los pases (Figura 4.1). La eficacia biológica media en el pase 1 era de $0,52 \pm 0,05$ (media \pm error estándar), de $0,79 \pm 0,04$ en el pase 11 (un 52% más que el inicial), de $1,03 \pm 0,09$ en el 21 (un 98% más) y, finalmente, esta subida experimentaba un gran incremento (de más de un 200%) en el pase 31 hasta llegar al valor de $1,57 \pm 0,25$. Así, al realizar un ANOVA con medidas repetidas, seguida de una comparación de los valores del pase 1 y pase 31 corregida por Bonferroni, este incremento era estadísticamente significativo con un $p\text{-valor} < 0,0001$.

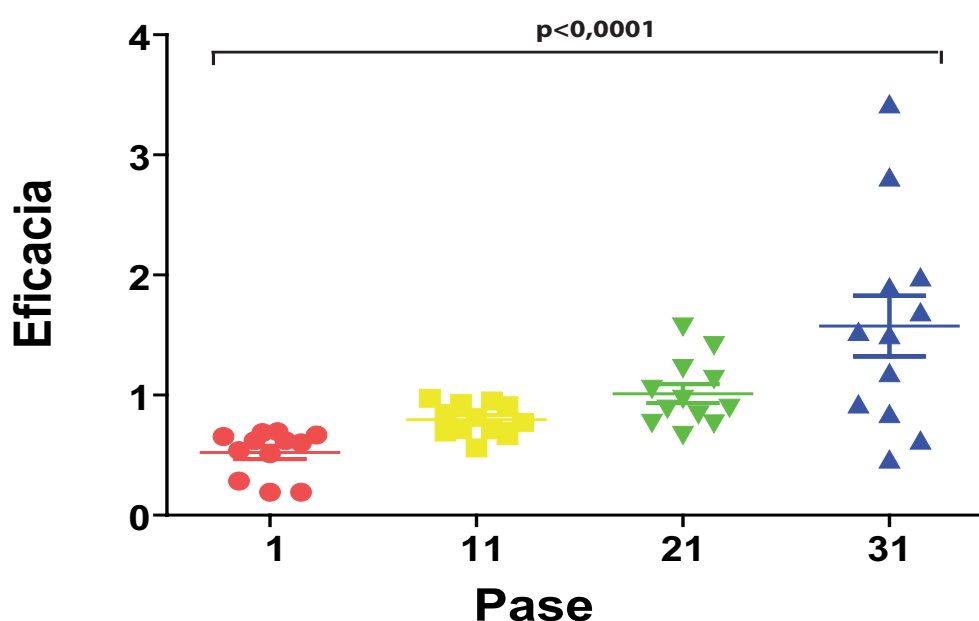


Figura 4.1 Valores de eficacia biológica a lo largo de pases seriados en gran tamaño poblacional de clones del VIH-1. Representación de los valores de eficacia en el pase 1 (●), pase 11 (■), pase 21 (▼) y pase 31 (▲). Se muestra la media de los valores por pase y su error estándar. Se muestra el p -valor de la comparación de medias entre el pase inicial y final.

A pesar de este aumento global, la dispersión de los valores, como se puede observar en la Figura 4.1, se incrementó a lo largo de los pases. Al examinar los virus individualmente se pudo detectar un comportamiento muy heterogéneo entre ellos (Figura 4.2). En primer lugar, aunque la mayoría de los virus presentaba un aumento considerable de su eficacia biológica en el pase 31, el camino seguido había sido muy distinto. Algunos virus mostraban un incremento continuo de eficacia, como son los virus D1, E1.5, G1, H1.5 e I1; otros virus, habiendo subido en el pase 11, se estancaron en el pase 21 y volvieron a subir en el pase 31, caso de los virus D2 y K1; finalmente, virus que sólo a partir del pase 21 o incluso del 31 habían sido capaces de incrementar su eficacia, como los virus I5 y G2.

Sin embargo, tres virus se apartaban de este patrón de incremento final de eficacia. En primer lugar, el virus K2 no había sido capaz de recuperar eficacia a lo largo de los 30 pases. En segundo lugar, los virus D1.5 y G1.5 sufrieron, sorprendentemente, una caída de sus valores de eficacia biológica en el pase 31, llegando incluso a valores al nivel del pase inicial en el caso de G1.5. Esto ocurría habiendo sido dos de los virus con mayor eficacia en el pase 21 (Figura 4.2).

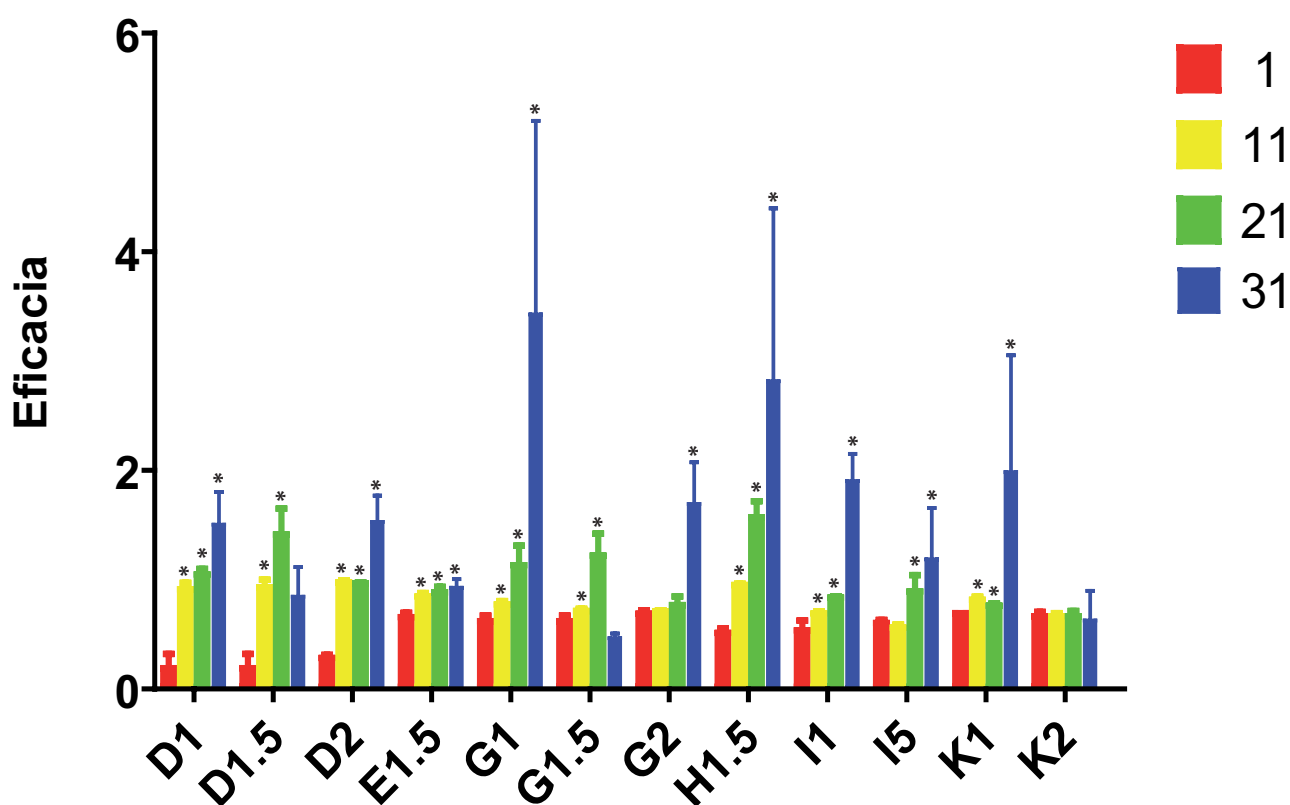


Figura 4.2 Valores de eficacia por virus y pase. Los valores de eficacia biológica se obtuvieron mediante ensayos de competición (explicado en materiales y métodos). Cada columna representa los valores de eficacia de cada virus en todos los pases (para cada pase se ha utilizado el color indicado en la leyenda). Los asteriscos (*) sobre las columnas indican que la diferencia entre el valor de eficacia del pase representado y el valor inicial es estadísticamente significativa (p -valor $<0,1$) utilizando una prueba de comparación de pendientes.

Para completar este análisis, se decidió comparar los valores de eficacia biológica de los virus recuperados, con sus valores iniciales antes de los pases en cuello de botella. Este contraste sólo se pudo realizar con cuatro linajes, D, G, I y K, ya que los valores de eficacia biológica de los virus E y H se habían calculado utilizando otra metodología de determinación de eficacia y, por lo tanto, no eran comparables. Como se puede comprobar en la Figura 4.3, en tres de los cuatro linajes al menos un virus ha sido capaz, no sólo de recuperar la eficacia inicial, sino de superarla. Todos los virus D y

los virus G1, G2 y K1 han sido capaces de sobrepasar los valores iniciales, mientras que el virus G1.5 la igualó en pase 21, pero a causa de su caída posterior ha vuelto a valores inferiores. El virus K2 al no haber sido capaz de recuperarse en todo el proceso, lógicamente no llegaba a los valores de su antecesor. Por último, el único linaje que no había llegado a la eficacia inicial, habiendo recuperado eficacia a lo largo de los pases de recuperación, era el linaje I, precisamente el de mayor valor inicial de todos los clones inicialmente aislados.

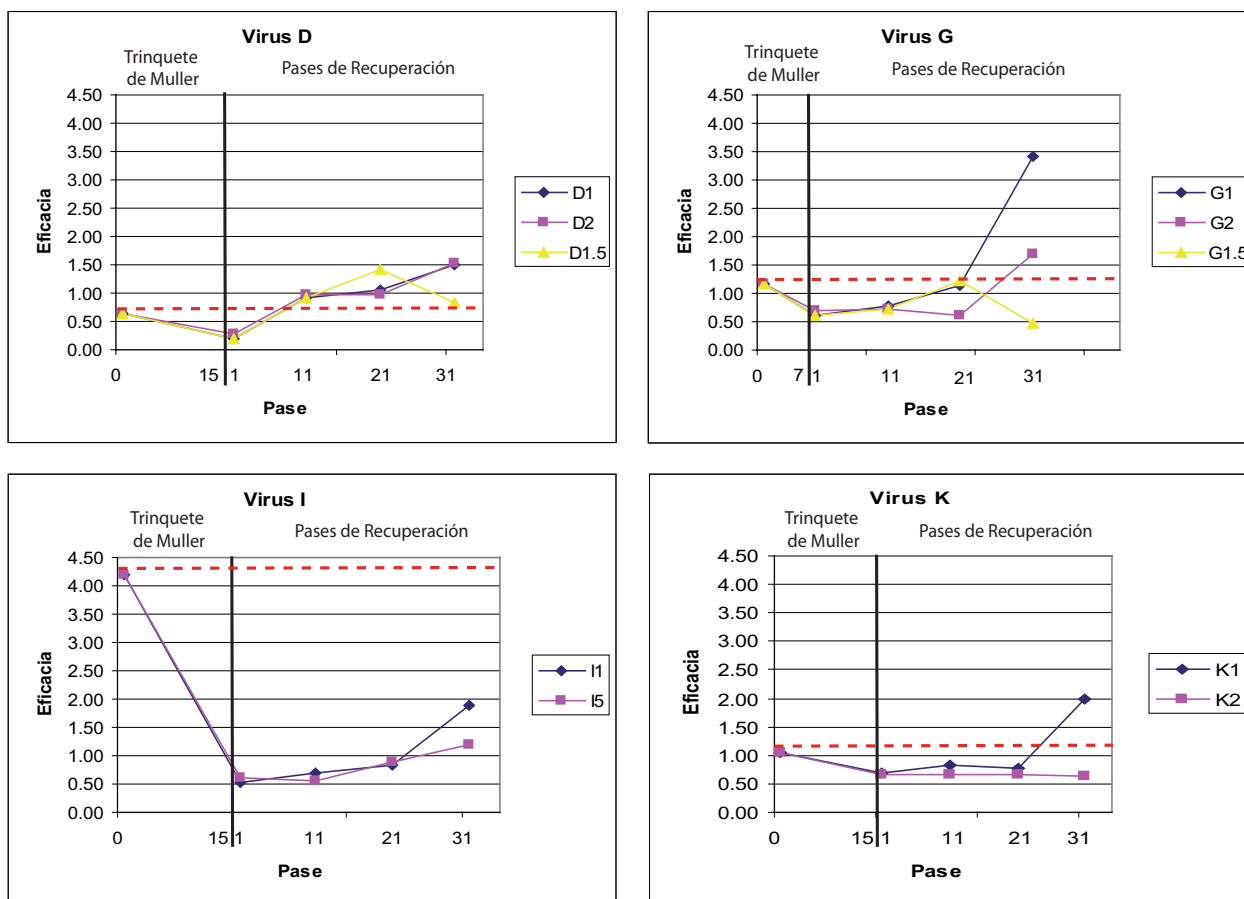


Figura 4.3 Valores de eficacia de los clones originales y de los virus después de los pases en cuello de botella y durante los pases en gran tamaño poblacional. Los primeros 15 pases representados corresponden a los pases placa a placa, excepto en el linaje G, que se extinguió en el pase 7. Los datos en estos pases son comunes a todos los virus de cada linaje, ya que los distintos clones luego estudiados se derivaron de las poblaciones finales de los cuellos de botella. Los pases placa a placa se separan mediante una línea vertical de los 30 pases posteriores en gran tamaño poblacional. Estos pases de recuperación se llevaron a cabo con distintos clones provenientes de la misma población. Los diferentes gráficos representan los distintos linajes estudiados y en las leyendas se muestra la representación de cada uno de los virus pertenecientes a estos linajes. La línea discontinua roja muestra el valor inicial de eficacia de cada clon original.

En resumen, el análisis de la eficacia biológica de los virus estudiados mostraba un aumento continuado de los valores globales, con un incremento muy significativo en el pase 31. Sin embargo, el comportamiento de los diferentes virus era muy heterogéneo. Por último, en la mayoría de los virus en los que se pudo comprobar (el 70%), la eficacia del clon original había sido igualada o superada.

4.2 Análisis genéticos de los virus estudiados.

4.2.1 Análisis de los genomas completos globales de los virus.

4.2.1.1 Estudio de las mutaciones aparecidas a lo largo del proceso de recuperación de eficacia biológica.

El primer paso para estudiar los factores responsables de esta recuperación de eficacia, fue analizar las mutaciones aparecidas en los genomas completos globales de los virus a lo largo de todo el proceso. Para ello, se utilizaron los genomas completos previamente obtenidos del pase 1 y 11 (Borderia y cols., 2010), secuenciando a su vez los genomas completos de cada virus en los pases 21 y 31.

Este examen mostró una acumulación muy baja de mutaciones, 24 en total, en los primeros 10 pases (pase 11). Incluso, cinco de los virus estudiados no presentaban ningún cambio en sus secuencias globales. Este panorama cambió en los siguientes 10 pases. Así, en el pase 21, ya se observaba un aumento exponencial en la acumulación de cambios, llegando a un número total de 95 mutaciones con respecto a los virus iniciales. Este incremento se estabilizó en los últimos 10 pases (pase 31), presentando el pase final 100 cambios totales con respecto a las semillas iniciales (Tabla 4.1).

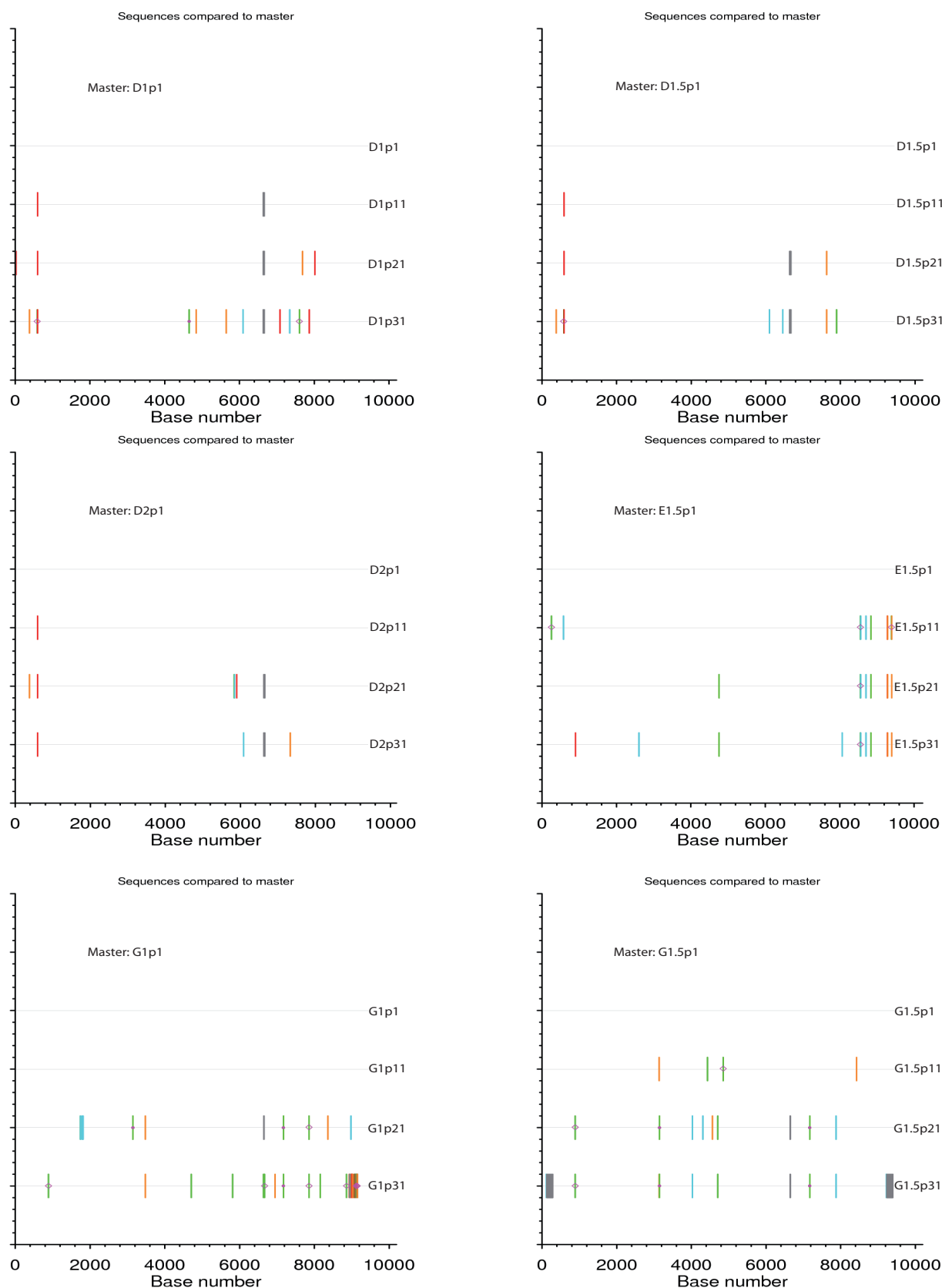
Mutaciones en pase 11				
Porcentaje	Nuevas Mutaciones	No Sinónimas	Sinónimas	No Codificantes (NC)
	24	19,0	2,0	5,0
%	100	79,2	8,3	20,8
% Sin NC		90,5	9,5	
Total	24			

Mutaciones en pase 21				
Porcentaje	Nuevas Mutaciones	No Sinónimas	Sinónimas	No Codificantes (NC)
	79	64,0	13,0	2,0
%	100	81,0	16,5	2,5
% Sin NC		83,1	16,9	
	Fijadas en P11			
	16	11,0	2,0	3,0
%	100	68,8	12,5	18,8
% Sin NC		84,6	15,4	
Total	95			

Mutaciones en pase 31				
Porcentaje	Nuevas Mutaciones	No Sinónimas	Sinónimas	No Codificantes (NC)
	62	46,0	7,0	8,0
%	100	74,2	11,3	12,9
% Sin NC		86,8	13,2	
	Fijadas en P21			
	37	32,0	2,0	3,0
%	100	86,5	5,4	8,1
% Sin NC		94,1	5,9	
Total	99			
	Fijadas en P21(sólo)			
	21	21,0	0,0	0,0
%	100	100,0	0,0	0,0
% Sin NC		100,0	0,0	
Total	121			
	Fijadas en P11			
	17	12,0	2,0	3,0
%	100	70,6	11,8	17,6
% Sin NC		85,7	14,3	
Total	100			

Tabla 4.1 Resumen de las mutaciones aparecidas en los diferentes pases de recuperación de eficacia biológica. Se indican las mutaciones nuevas encontradas y las que permanecen fijadas de pases anteriores. Además se indica el número y porcentaje de mutaciones sinónimas, no sinónimas y las que se encuentran en regiones no codificantes.

Utilizando la aplicación highlighter de la base de datos de Los Alamos (<http://www.hiv.lanl.gov/content/sequence/HIGHLIGHT/highlighter.html>), se generaron representaciones de la acumulación de cambios en los genomas completos. Como se puede observar en la Figura 4.4, dicha acumulación se producía en cada linaje en diferentes partes del genoma e incluso el patrón cambiaba durante los pases. Esto indicaba que los virus habían seguido diferentes caminos evolutivos en la recuperación de eficacia.



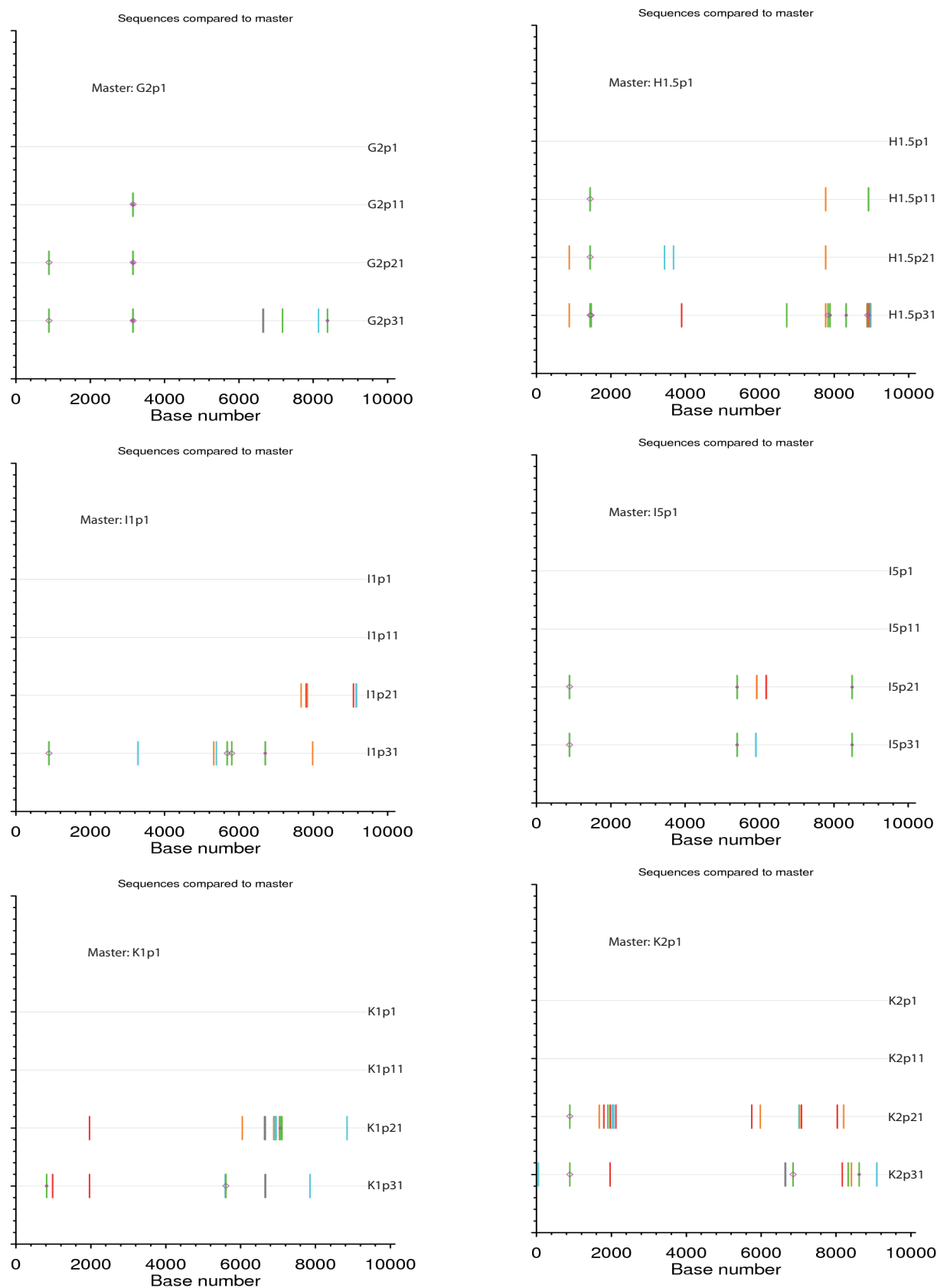


Figura 4.4 Mutaciones aparecidas en los genomas completos de los diferentes virus a lo largo de los pases de recuperación de eficacia. En el eje de abscisas se representa el número de base dentro del genoma completo del VIH. El código de colores utilizado fue verde para A, rojo para T, naranja para G, azul claro para C, azul oscuro para otros códigos de la IUPAC y gris para las deletiones. Además se señalan los cambios asociados a APOBEC con círculos (no aplicables en nuestro caso) y los cambios G-A con un rombo. Los cambios se indicaron con respecto al virus inicial de los pases de recuperación.

En cada pase, más del 80% de las mutaciones eran no sinónimas (Tabla 4.1 y Figura 4.5A), apareciendo también un porcentaje relativamente alto de mutaciones en zonas no codificantes, en concreto en la región LTR (13% del total de las mutaciones fijadas en el pase 31).

Además, aparecían deleciones en varios sitios del genoma en los diferentes virus. A destacar son las deleciones aparecidas en la región V1-V2 del gen *env*, que afloraron en la mayoría de los virus y de las que se hablará en un apartado posterior. También se detectó una gran deleción que surgió en el virus G1.5, que cayó de eficacia en los últimos pases (Figura 4.4). Esta deleción se situaba en la región U3 de la 5'LTR y, en el extremo 3', también en el gen *nef*.

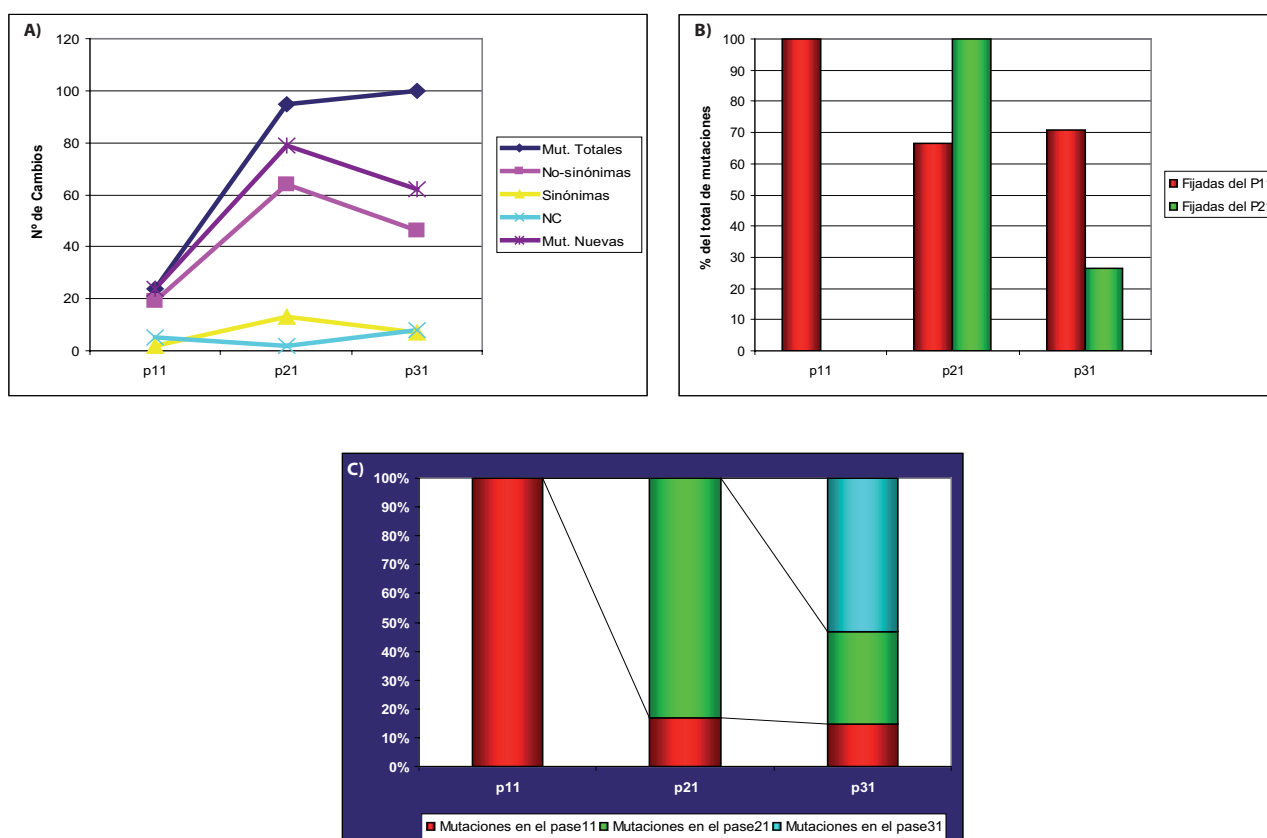


Figura 4.5 Dinámica de aparición de mutaciones durante la recuperación de la eficacia biológica. A) Número de cambios aparecidos en los pases de recuperación tanto totales, nuevas, sinónimas, no sinónimas y en regiones no codificantes. B) Porcentaje del total de mutaciones aparecidas en el pase 11 (rojo) y 21 (verde) que se mantienen en los subsiguientes pases. C) Porcentaje del total de cambios en cada pase con respecto al pase inicial correspondiente a mutaciones aparecidas en pases anteriores.

Estudiando en detalle el proceso de aparición y fijación de los cambios en las secuencias globales, se pudo comprobar que alrededor del 70% de las mutaciones presentes en el pase 11 se mantuvieron en los pases subsiguientes, mientras que sólo el 26% de las aparecidas en el pase 21 se conservaron en el 31 (Figura 4.5B). Además, las mutaciones que no se mantuvieron en los pases 21 y 31, revirtieron al nucleótido original. Estos datos estaban indicando la importancia de las primeras mutaciones fijadas en la recuperación de eficacia biológica. Por otro lado, las mutaciones aparecidas en los pases más avanzados, debido a que en general no se mantenían en las secuencias globales a lo largo de los pases, estarían reflejando la complejidad de la población más que la importancia de las mutaciones individuales.

También se analizó el porcentaje que suponían las mutaciones aparecidas en pases anteriores respecto del total de cambios en los pases 21 y 31 (Figura 4.5C). Así, se pudo advertir que las mutaciones adquiridas en pases previos pasaron de suponer cerca del 20% en el pase 21 a alrededor del 50% en el pase 31. Por su parte, la acumulación de mutaciones nuevas, que pasó de 24 en el pase 11 a 79 en el pase 21, bajó hasta las 62 en el último pase. Todos estos datos, sugieren un mantenimiento de los cambios genéticos iniciales, a partir de los cuales poder explorar el paisaje evolutivo, y una estabilización de las mutaciones acumuladas.

Con respecto a la distribución de las mutaciones a lo largo del genoma, se pudo observar un cambio continuo en los genes que acumulaban estos cambios (Figura 4.6A). En los primeros 10 pases, el mayor porcentaje de acumulación total lo tuvo el gen *pol*, mientras que en los 10 siguientes, se acumularon sobre todo en *gag* y *env*, acabando en el pase 31 con una preponderancia de *env*. Siguiendo esta línea, se procedió a analizar la acumulación de mutaciones en cada gen, esta vez normalizando con respecto a su tamaño. Así, se podía tener una idea real de los cambios que se acumulaban con respecto a los que cabría esperar por el tamaño de cada gen si se produjesen al azar (Figura 4.6B). En los primeros pases se pudo ver una sobre-representación de mutaciones en LTR, *rev* y *nef*; en el pase 21 este patrón cambió a *gag*, *tat*, y *env* y en el pase final LTR, *vif*, *vpr*, *vpu* y *env*. En cada uno de los pases, la sobre-representación de mutaciones en los genes estructurales era común a la mayoría de los virus, mientras que en los genes accesorios era debida a la acumulación de mutaciones en alguno de los genomas. También es reseñable la disminución de mutaciones acumuladas que ha sufrido el gen *pol* a lo largo de los pases, resultado lógico debido al efecto deletéreo que la mayoría de las mutaciones pueden tener en este gen.

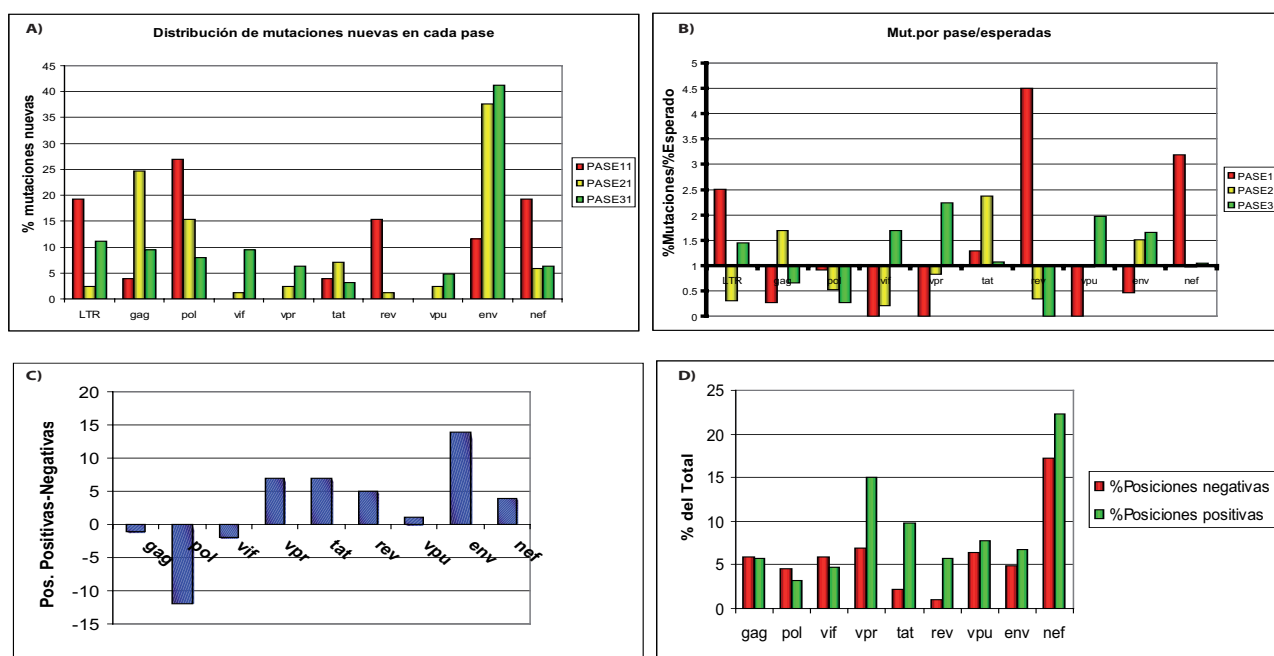


Figura 4.6 Distribución de mutaciones en los diferentes genes y presión selectiva en los genomas consenso. A) Porcentaje de mutaciones nuevas aparecidas en los diferentes genes en el pase 11 (rojo), 21 (amarillo) y 31 (verde). B) Porcentaje de mutaciones nuevas respecto al porcentaje esperado por el tamaño del gen aparecidas en los diferentes genes a lo largo de los pases de recuperación (colores por pase igual que en A). C) Diferencia entre las posiciones significativamente seleccionadas positivamente y las seleccionadas negativamente en cada gen durante todo el proceso de recuperación. D) Porcentaje de posiciones sobre el total de cada gen positivamente (verde) y negativamente (rojo) seleccionadas durante los pases en gran tamaño poblacional.

Para contrastar la significación de la presión selectiva que estaba operando en el proceso, se realizó un contraste de selección por codones en cada gen, utilizando los genomas completos de todos los virus desde el pase 1. En este análisis se observó cómo las posiciones bajo selección positiva preponderaban sobre todo en el gen *env*, con un 40% más de sitios seleccionados positivamente que negativamente, resultado que también se daba en *vpr*, *tat*, *rev* y *nef* (Figura 4.6C). Por su parte, las posiciones seleccionadas negativamente predominaban en *pol*. En cuanto al porcentaje de posiciones que se separaban significativamente de la neutralidad con respecto al total de posiciones que constituían cada gen, cabe destacar que se mantenían en torno al 5-10% de las posiciones totales en casi todos los genes, excepto en *vpr* y *nef* (Figura 4.6D). En estos dos genes el porcentaje de posiciones que estaban sometidas significativamente a presión selectiva, subía hasta más del 15% del total. Este cambio continuo de la localización de las mutaciones sugería un ambiente selectivo cambiante, mientras que la presión positiva que había operado en la mayor parte del genoma, indicaba la presencia de una selección constante.

A pesar de lo heterogéneo de la fijación de mutaciones a lo largo del genoma, sí que se pudieron observar eventos de convergencia entre los diferentes virus. En concreto, hubo 9 eventos de convergencia, 8 aparecieron en diferentes virus del mismo linaje, mientras que uno apareció en seis virus de tres linajes diferentes. Las mutaciones convergentes fueron las siguientes:

- El cambio G593T de la LTR apareció en todos los virus del linaje D desde el pase 11, además de haber aparecido ya en un trabajo previo (Yuste y cols., 2005).

- La mutación A379G, también en la LTR es una reversión de un cambio aparecido en el linaje D durante los pases seriados placa a placa, que se produce en los virus D1p31, D1.5p31 y D2p21, aunque en D2p31 vuelve a aparecer en mezcla. Esta mutación se encuentra en el sitio de unión a SP1-III, factor de transcripción de gran importancia para la replicación viral, de la región U3 (McAllister y cols., 2000). Probablemente la presencia de la G era perjudicial para estos virus ya que se adquirió en el proceso de Trinquete de Muller, donde se acumulan mutaciones deletéreas.

- Otra reversión en la LTR, la G586A, ocurre también en D1p31 y D1.5p31 devolviendo el nucleótido inicial en estos virus que habían fijado la G al hacer la semilla después del Trinquete de Muller. Probablemente, esta mutación permite a los virus recuperar eficacia, pero es ineficaz a largo plazo, revirtiéndose posteriormente por ese motivo.

- La mutación V35I de la proteína p17 fue la que apareció en los seis virus de tres linajes diferentes (G, I y K). Es importante recalcar que ninguno de estos virus había acumulado mutaciones comunes en los pases placa a placa, luego no se podía achacar a este cambio un efecto compensatorio de una mutación previa. Esta mutación estaba apareciendo probablemente debido a su efecto beneficioso en eficacia. La mencionada posición ha sido relacionada con la recuperación de eficacia de mutantes resistentes a inhibidores de proteasa (Gatanaga y cols., 2002). El mencionado cambio se localiza en el extremo C terminal de una α -hélice, que corresponde con la señal de localización nuclear (NLS, en inglés) (Bukrinsky y cols., 1993).

- Otro cambio convergente, la E196K, se localizaba en la RT. Apareció en todos los clones G y podría tener importantes efectos debido a su proximidad al centro catalítico de la enzima situado en los residuos 185 y 186 (Jacobo-Molina y cols., 1993) y al cambio drástico de aminoácido que suponía.

- En los virus G1p31 y G1.5p31 se fijó la mutación L158I en la integrasa, aunque en el G1.5 ya aparecía en el pase 21. Este cambio se situaba en la región catalítica, cerca del tercer residuo de la tríada catalítica (E152) (Pommier, Johnson y Marchand, 2005), por lo que es probable que tenga importancia en la función de esta proteína.

- También en dos de los clones G apareció el cambio S314K en la región V3 de la glicoproteína gp120, que es precisamente una de las mutaciones asociadas con el cambio de fenotipo R5/X4 (De Jong y cols., 1992).

- El cambio L16F en la proteína p7 apareció en los dos virus del linaje K, esta mutación no tiene descrito en la literatura ningún efecto significativo.

- Finalmente, la mutación V8A en la proteína Vpu apareció en los virus D1p31 y D1.5p31. Su localización se encuentra dentro del dominio transmembrana de la proteína, que juega un importante papel en la función de la misma.

Estas mutaciones convergentes nos dan más información sobre el proceso seguido por los virus para recuperar eficacia. En primer lugar, la mayoría de las mutaciones parecen estar asociadas con importantes efectos fenotípicos. La presencia de una mutación común en tres linajes, nos indica la gran importancia de la misma. Por otro lado, la mayor parte de las convergencias se dan entre réplicas del mismo virus pasadas en paralelo (D1-D1.5 y G1-G1.5), indicando los pasos que esos virus tienen que seguir obligatoriamente para recuperar eficacia.

Las reversiones que aparecen únicamente se dan tras 30 pases, lo que indicaría que en primer lugar esos cambios se han intentado compensar y, una vez generada la variabilidad suficiente, se han conseguido revertir.

Por último, es de reseñar que estos cambios con importantes efectos fenotípicos han surgido sin la presencia de ninguna fuerza selectiva externa (antivirales, sistema inmune...), sino por el simple cultivo en poblaciones de gran tamaño celular.

4.2.1.2 Aumento de las posiciones heterogéneas a lo largo de los pases de recuperación.

Además del aumento tan significativo de las mutaciones fijadas en las secuencias globales a lo largo de la recuperación, también se pudo detectar un aumento espectacular en las posiciones que presentaban mezcla de nucleótidos en la secuencia global. En estas posiciones siempre se observó mezcla de dos nucleótidos, excepto en una posición de H1.5p31 en la que se observaba la mezcla de tres, y siempre una de las bases presentes era la original.

En los virus iniciales y en los correspondientes al pase 11, no se observó ninguna posición heterogénea. En los siguientes pases aparecieron multitud de dichas posiciones, observándose 173 en el pase 21 y 217 en el 31. Al realizar un análisis de correlación sencillo, para comprobar la relación entre la aparición de estos sitios heterogéneos y la eficacia observada en los virus, se obtuvo una correlación estadísticamente significativa con un *p-valor* de 0.0047 en una prueba de Pearson. Aunque esta prueba estaba algo sesgada por los valores negativos de sitios heterogéneos de los pases 1 y 11, parecía indicar que el aumento de la heterogeneidad en las poblaciones que estas posiciones mostraban, estaba relacionado con el aumento de la eficacia biológica observado.

4.2.1.3 Aparición de deleciones en la región V1-V2 del gen *env*.

Como se ha indicado anteriormente, a lo largo de la recuperación de eficacia de los virus estudiados, se fueron fijando diversas deleciones en la zona variable V1-V2 de la proteína gp120. Al inicio de la recuperación todos los virus, excepto E1.5 y H1.5, tenían una longitud de 218 nucleótidos en dicha región, igual a la del aislado original s61. A lo largo de los pases, los diferentes virus han ido adquiriendo deleciones en la región (Figura 4.7). Al final de los 30 pases, todos los virus presentaban deleciones en V1-V2, excepto el I1; además, las deleciones que presentaba el otro virus de su linaje, I5, eran muy minoritarias (con una frecuencia en la población menor del 3%).

El tamaño y número de estas deleciones era muy diverso. Se observaron deleciones con tamaños que iban desde 6 nucleótidos (nt) (virus K2) hasta 45 (D1.5 y D2), imponiéndose en algunos casos siempre los virus con un tamaño (deleción de 33 nt en los virus del linaje D o de 18 en los virus E1.5 y H1.5) y siendo el tamaño fijado muy diferente en otros casos (linaje G o K). Por su parte, la cantidad de variantes con deleciones aparecidas también era muy diferente, yendo desde un solo variante en E1.5 y H1.5, hasta 6 variantes diferentes en K2. En la mayoría de los virus estas deleciones, una vez aparecidas, se han hecho dominantes en la cuasiespecie, excepto los virus G1

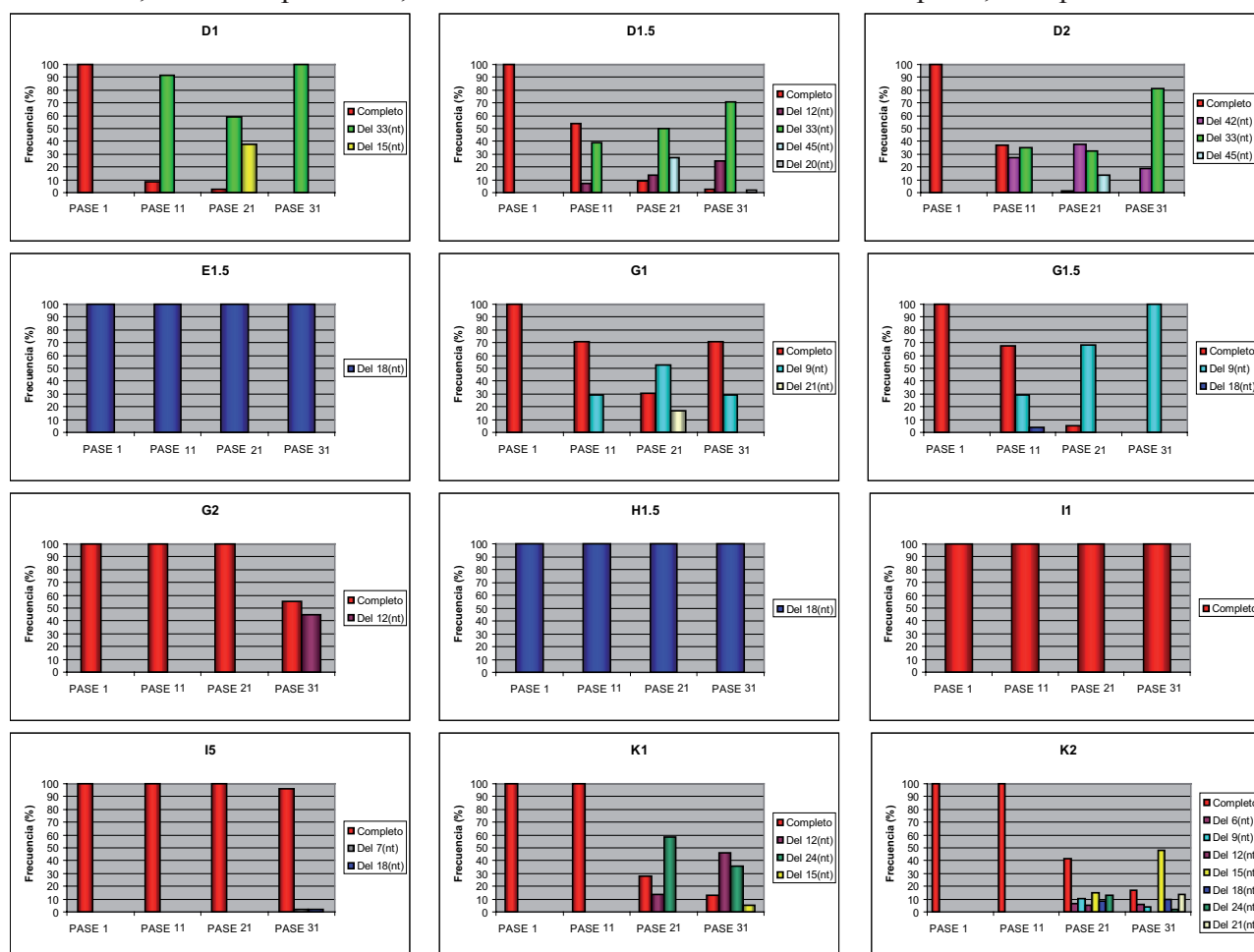


Figura 4.7 Deleciones en la región V1/V2 del gen *env* de los virus. Aparición de genomas que presentan deleciones en la mencionada región en los pases 11, 21 y 31, representando su frecuencia en las cuasiespecies virales. Esta frecuencia fue calculada utilizando GeneScan (ver Materiales y Métodos). Los genomas completos se representan en rojo, mientras que los genomas delecionados siguen el código de colores mostrado en las leyendas según el tamaño de la deleción que presentan.

y G2 que, teniendo deleciones de manera significativa en la población, aun presentaban el virus completo como el más frecuente. Por lo tanto, la aparición de estas deleciones en la región V1-V2 del gen *env* parece haber jugado un papel importante y beneficioso en el proceso de recuperación de eficacia biológica.

4.2.2 Análisis de las cuasiespecies virales.

4.2.2.1 Aumento de la heterogeneidad de las cuasiespecies a lo largo de los pases de recuperación.

Debido a la falta de un patrón común en las mutaciones fijadas en las secuencias globales, a la falta de relación aparente entre los cambios adquiridos y la recuperación de eficacia biológica acumulada, al aumento de posiciones heterogéneas en los genomas globales, que indicaban un aumento de la heterogeneidad de la población, y a los incrementos de eficacia observados en el pase 11 sin modificación de las secuencias globales (Borderia y cols., 2010), se decidió llevar a cabo un análisis más profundo y examinar la composición de las cuasiespecies virales que conformaban los diferentes virus estudiados. Para ello, se obtuvieron clones de cada uno de los virus en 4 regiones genómicas diferentes, que abarcaban LTR-*gag*, *vpu*, *env* (V1-V2) y *env* (V3-V4), y se analizaron sus secuencias. Estas secuencias se utilizaron para medir la heterogeneidad de las cuasiespecies por medio de métodos de Máxima Verosimilitud.

En primer lugar se obtuvo el modelo evolutivo que seguían los virus en las distintas regiones estudiadas y utilizando el valor medio ponderado de los parámetros evolutivos estimados, se calcularon los valores de heterogeneidad por Máxima Verosimilitud (ver Materiales y Métodos apartado 3.4.2.5). Una vez obtenidos, se probó que no hubiera diferencias significativas entre las regiones de un mismo pase utilizando para ello una prueba ANOVA. No se encontraron diferencias entre los valores de cada región en el mismo pase, por lo que se pudo utilizar la media de los valores por cada virus y pase para hacer los análisis subsiguientes.

En consecuencia, se procedió a examinar la heterogeneidad global a lo largo de los pases, observándose un incremento de los valores que fueron desde $3,273 (\pm 0,209) \times 10^{-3}$ sustituciones por sitio (s/s) en el pase 1, hasta $3,760 (\pm 0,256) \times 10^{-3}$ s/s en el pase 11, $4,342 (\pm 0,344) \times 10^{-3}$ s/s en el pase 21 y $4,442 (\pm 0,400) \times 10^{-3}$ s/s en el pase 31 (Figura 4.8). Al probar la significación de estas diferencias, se obtuvo un incremento significativo tanto en el pase 21 como en el 31 con respecto al pase inicial ($p\text{-valor} < 0,05$) una vez realizado un ANOVA y corregido los p-valores de las comparaciones dos a dos utilizando la corrección de Bonferroni. Este resultado demostró el aumento continuado de la complejidad de las cuasiespecies virales durante los pases de recuperación.

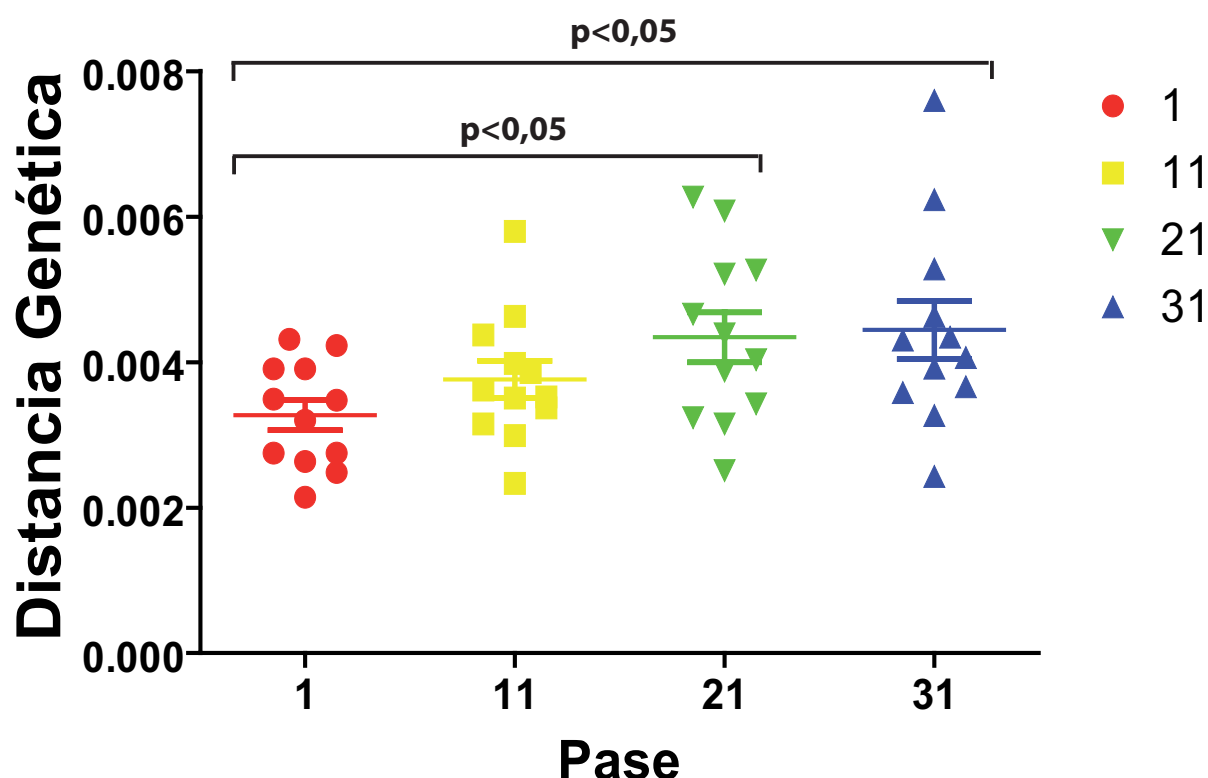


Figura 4.8 Heterogeneidad de las cuasiespecies virales a lo largo de los pases de recuperación. La heterogeneidad de las cuasiespecies fue medida como la distancia genética media (número de sustituciones por sitio) como se explicó en Materiales y Métodos. La heterogeneidad dentro de las cuasiespecies se analizó utilizando los valores medios de cada virus en las cuatro regiones examinadas. Se realizó una comparación entre pases y los *p*-valores significativos (después de haber sido corregidos por el método Bonferroni) se muestran en la figura encima de los pases comparados. Los valores de cada virus y las medias y errores estándar por pase se muestran siguiendo el mismo código que en la Figura 4.1.

4.2.2.2 Existencia de una correlación global significativa entre heterogeneidad y eficacia biológica a lo largo de los pases de recuperación.

Con los resultados obtenidos anteriormente, se decidió estudiar si el aumento de eficacia estaba relacionado con el aumento de heterogeneidad observado, ya que los patrones de variación de ambas variables durante los pases de recuperación eran muy similares. Para ello se realizaron diversas pruebas de correlación. En primer lugar se ensayó la correlación tanto con un método paramétrico (Pearson), ya que se disponía de valores suficientes para asumir una distribución normal, como con un método no paramétrico (Spearman) utilizado en los trabajos de los pases anteriores (Borderia y cols., 2010; Lorenzo-Redondo, Borderia y Lopez-Galindez, 2011). Ambos métodos fueron utilizados agrupando todos los valores de heterogeneidad y eficacia de todos los pases analizados en este trabajo. El resultado obtenido fue una correlación estadísticamente significativa entre eficacia y heterogeneidad independientemente del método utilizado, con *p*-valor de 0,0338 utilizando Pearson y 0,0057 utilizando Spearman (Figura 4.9).

A pesar de esta fuerte correlación, ciertos virus como D1.5p21, G1.5p21 o G1p31, se salían de este patrón en los diferentes pases. Estos virus presentaban unas de las eficacias más altas en sus respectivos pases, pero baja heterogeneidad. Debido a los valores exhibidos por estos virus, se observaban diferentes modelos de correlación si se analizaba cada pase por separado, con una

correlación más fuerte en el pase 11, que en el 21 y en el 31. Por este motivo y para confirmar la correlación global previamente obtenida, se decidió probar modelos de regresión de efectos mixtos considerando los pases como medidas repetidas y, de este modo, tener en cuenta la correlación entre pases. El mejor modelo obtenido por máxima verosimilitud siguiendo diversos criterios de decisión, pero principalmente el AIC, fue un modelo que incluía una matriz de correlación sin estructura de los efectos repetidos y consideraba los efectos aleatorios entre sujetos. Este modelo le otorgaba una correlación positiva a la heterogeneidad con respecto a la eficacia con un *p*-valor de 0,07. Aunque este valor no era significativo al 95%, si lo era al 90%, lo que resulta aceptable en un modelo tan restrictivo como el utilizado. Además, otros modelos con menor verosimilitud otorgaban valores de *p*-valor ya estadísticamente significativos al 95%.

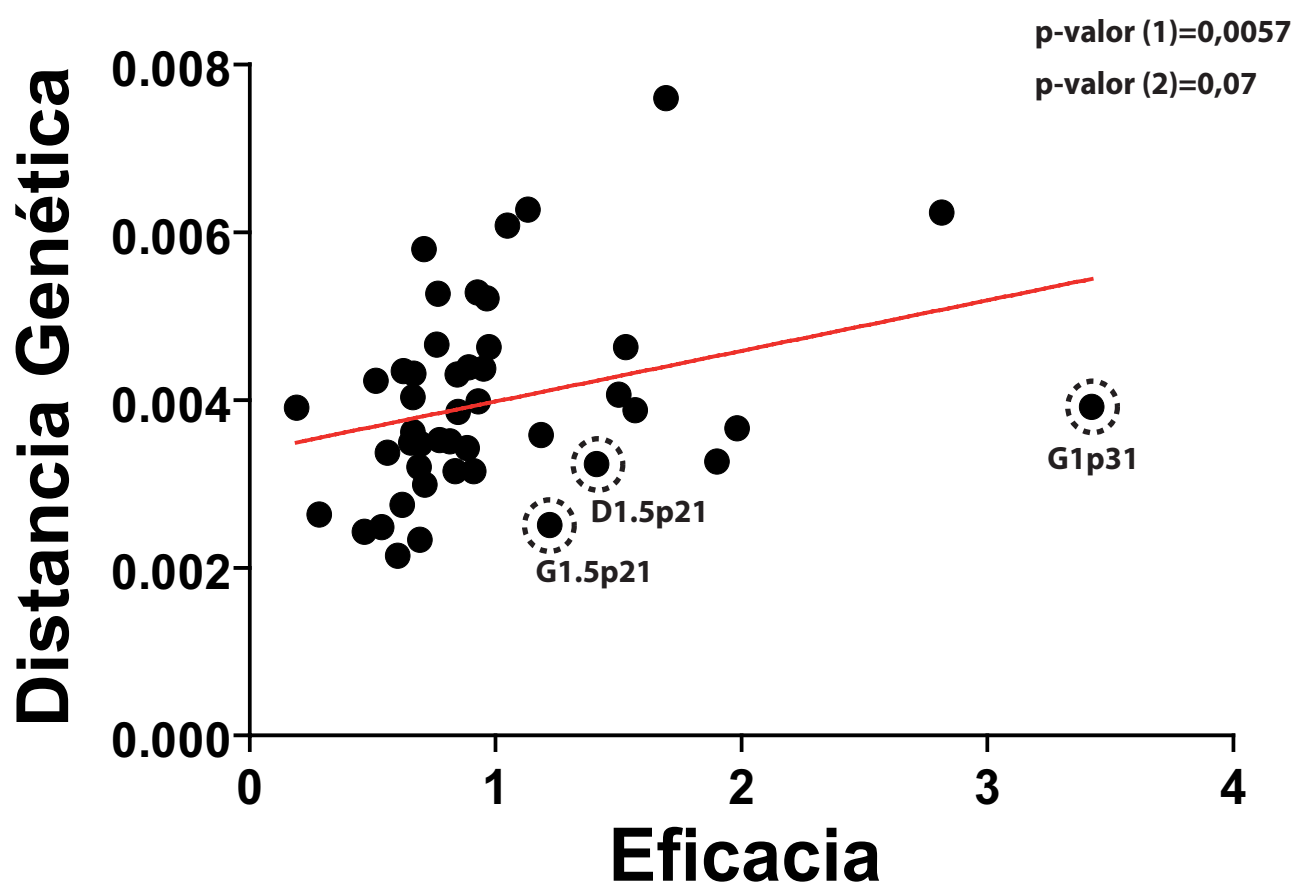


Figura 4.9 Análisis de correlación entre la heterogeneidad de las cuasiespecies y la eficacia biológica. La correlación global a lo largo de los pases de recuperación se obtuvo utilizando todos los valores de heterogeneidad (medida como distancia genética) y eficacia de los virus en los pases 1, 11, 21 y 31. Se muestran los p-valores obtenidos mediante correlación simple de Spearman (1) y modelos lineales de efectos mixtos (2), que se utilizaron para analizar la significancia de la correlación estadística entre estas dos variables. Se indican los puntos que representan a los virus que se salieron del patrón general de sus respectivos pases.

En definitiva, se pudo demostrar que existe una correlación estadísticamente significativa entre la ganancia de eficacia y el aumento de complejidad de las cuasiespecies virales en el proceso de recuperación.

4.2.2.3 Análisis de las fuerzas selectivas operando en las cuasiespecies durante la recuperación de eficacia biológica.

A continuación, se decidió analizar más exhaustivamente la composición de las cuasiespecies virales y examinar la clase de mutaciones que presentaba la nube de mutantes. Para ello se calculó la cantidad de cambios sinónimos y no sinónimos dentro de la cuasiespecie con respecto al pase inicial. Este examen se hizo en cada una de las regiones en las que previamente se había analizado la heterogeneidad, no aplicándose a la región LTR al no ser codificante. El objetivo final de todo este estudio era conocer las fuerzas selectivas que estaban operando en el proceso.

En primer lugar, el cálculo de dS mostró un aumento del valor medio en el pase 11, que se mantuvo sin variación en los subsiguientes pases (Figura 4.10). Por otro lado, los valores de dN experimentaron un crecimiento exponencial a lo largo de los pases. Por último, se calculó el valor de dN-dS para todos los virus en todas las regiones por pase para probar estadísticamente la preponderancia de un tipo u otro de mutación. El resultado obtenido fue un cambio en la composición de la heterogeneidad de las cuasiespecies, pasando de una preeminencia de dS en los primeros pases, a un dominio de las dN cada vez mayor en los pases 21 y 31. Así, la diferencia entre dN y dS pasó de $-6,5 (\pm 3) \times 10^{-4}$ en el pase 11, a $4,6 (\pm 4) \times 10^{-4}$ en el pase 21 y $15 (\pm 5) \times 10^{-4}$ en el pase 31.

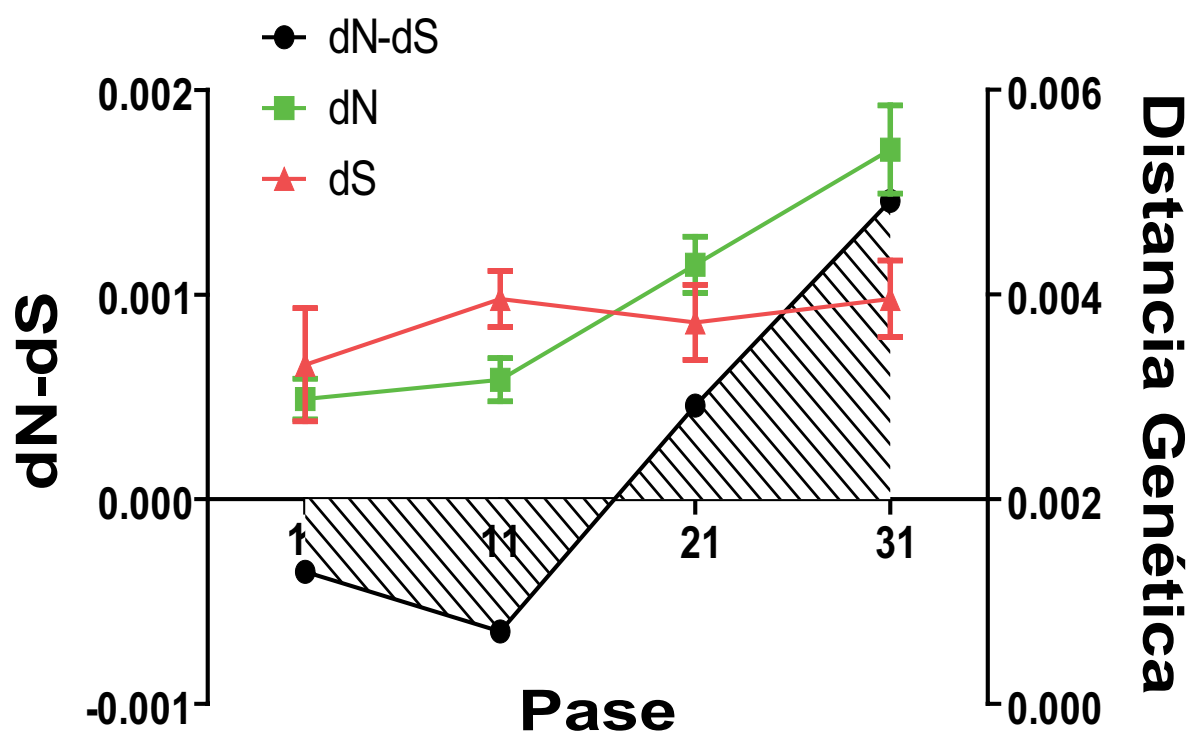


Figura 4.10 Análisis de mutaciones sinónimas y no sinónimas dentro de las cuasiespecies a lo largo de los pases de recuperación. Se calcularon las dN (número de sustituciones no sinónimas por posible sitio no sinónimo) y dS (número de sustituciones sinónimas por posible sitio sinónimo) dentro de las cuasiespecies de todos los virus en cada pase usando el método modificado de Nei-Gojobori. Para estimar los valores se asumió el modelo evolutivo de Jukes-Cantor. En el eje izquierdo se representa el valor medio por pase de dN-dS (área rayada) y en el eje de la derecha los valores medios con su error estándar por pase de dN (verde) y dS (rojo).

La fuerte preponderancia de las dN al final de los pases indicaba una fuerte selección positiva operando en el proceso. Al no coincidir el aumento exponencial de dN dentro de las cuasiespecies con un aumento en el número de mutaciones no sinónimas en las secuencias globales, que pasó de 64 nuevas en el pase 21 a 46 en el pase 31 (Tabla 4.1), este resultado indicaba que la composición de las cuasiespecies, al igual que las secuencias globales, estaba siendo moldeada por la selección positiva.

Finalmente, se llevó a cabo otro análisis de correlación para comprobar la relación entre el tipo de mutaciones en el espectro de mutantes y los valores de eficacia biológica. Como se puede ver en la Figura 4.11, había una correlación muy significativa entre los valores de dN y eficacia, mientras que esto no sucedía en el caso de dS. Lógicamente, los valores de dN-dS también correlacionaban significativamente con los de eficacia.

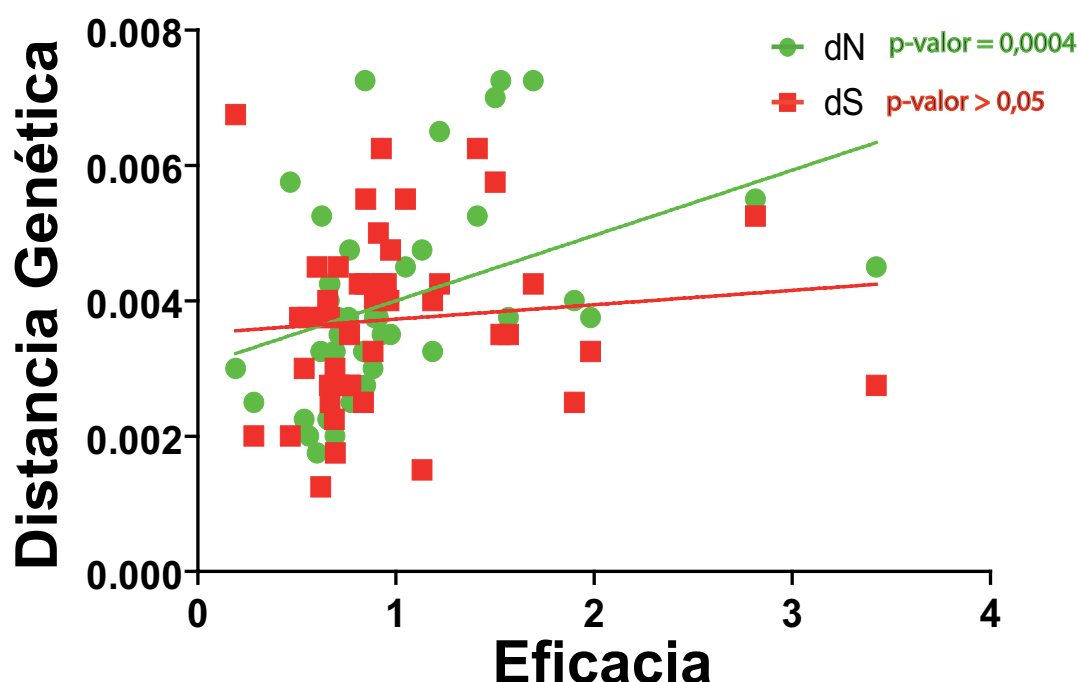


Figura 4.11 Correlación estadística entre la composición de dN y dS dentro de las cuasiespecies y los valores de eficacia biológica. Se analizó mediante correlación de Spearman la asociación entre los valores de dN (verde) y dS (rojo) con los valores de eficacia de cada virus en los diferentes pases. Se muestra el p-valor obtenido en los dos análisis, en los que solo se encontró asociación entre dN y eficacia biológica.

Por lo tanto, la selección positiva estaba dirigiendo mayoritariamente la evolución del proceso de recuperación, estando muy relacionada con el incremento de eficacia observada en los virus estudiados.

4.3 Estudios filogenéticos.

Con las secuencias obtenidas en los análisis genéticos, se llevaron a cabo diversos estudios filogenéticos con el fin de comprender mejor la evolución de los virus analizados en su camino hacia la recuperación de eficacia biológica. Estos estudios se realizaron tanto a nivel de secuencias globales, como a nivel de cuasiespecies.

4.3.1 Modelo evolutivo seguido por los virus en los pases en gran tamaño poblacional.

En primer lugar y al igual que se hizo con las secuencias de las cuasiespecies para calcular su heterogeneidad, se estimó el modelo evolutivo que mejor se ajustaba a las secuencias de los genomas completos globales de todos los virus estudiados durante la recuperación de eficacia. Para este propósito se utilizó el programa jModelTest en el que se calcularon los valores de verosimilitud de 24 modelos evolutivos y se seleccionó el que mejor se ajustaba a los datos siguiendo el criterio de información de Akaike. El mejor modelo obtenido fue el GTR con proporción de sitios invariables (I) y variación de las tasas entre sitios (G) con un peso de 0,9824, con lo que este modelo explicaba casi por completo los datos. Los virus presentan un 29,5% de sitios invariables y un valor de α de 0,011, que indicaba una clara distribución gamma de las tasas entre diferentes sitios. Este modelo elegido fue utilizado en análisis posteriores.

4.3.2 Estimación de la filogenia de los virus en el proceso de recuperación por Máxima Verosimilitud.

Con los parámetros estimados anteriormente, ya se estaba en disposición de evaluar la filogenia de los virus estudiados. Con este objetivo, se utilizó el programa PhyML para obtener un árbol filogenético de los citados virus por el método de Máxima Verosimilitud. Para realizar esta estimación se hizo uso de los parámetros evolutivos obtenidos en el programa jModelTest. En este caso, también se decidió utilizar las medias estimadas de los parámetros teniendo en cuenta los de todos los modelos probados. De todos modos, las medias eran casi iguales a los parámetros del modelo GTR+I+G, ya que su peso era muy grande. Utilizando esta metodología se obtuvo el árbol de Máxima Verosimilitud de los virus estudiados (Figura 4.12).

Este árbol mostraba una perfecta separación no sólo de cada linaje sino también de cada virus proveniente de los diferentes clones utilizados de cada linaje. Esto descartaba en primer lugar cualquier contaminación e indicaba la distinta evolución seguida por cada uno de ellos. El linaje I era el más próximo al virus original, mientras que D y K eran los más alejados. Como se puede observar, los virus han ido evolucionando continuamente, viéndose algunos saltos muy grandes en el pase 31 como los de G1p31 y H1.5p31, que coinciden con fuertes subidas de eficacia. La evolución de alguno de los virus ha sido siempre acumulativa, como son los casos de E1.5 y H1.5, mostrando un patrón temporal de acumulación de mutaciones a lo largo de los pases de recuperación. Mientras tanto, otros virus han seguido una evolución menos regular, casos de I5 o K1, en los que se han probado diversos

caminos de recuperación de eficacia. Esto último se puede comprobar por la existencia de ramas significativas externas a los virus de pases anteriores, lo que quiere decir que han fijado mutaciones en algún pase que luego se han visto reemplazadas otra vez por las originales (como ejemplo claro véase el virus K1).

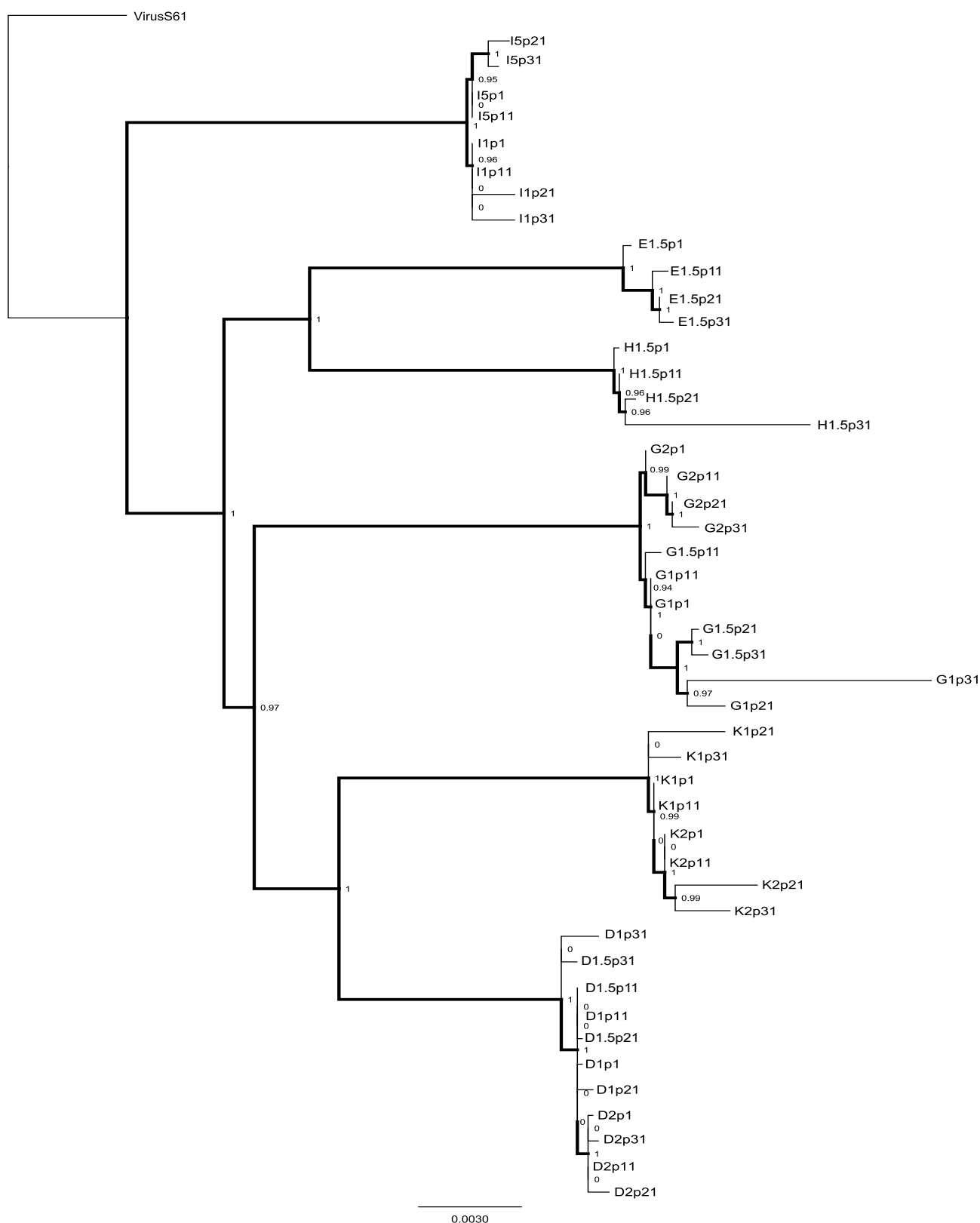


Figura 4.12 Árbol filogenético obtenido por Máxima Verosimilitud de los genomas completos de los virus en el proceso de recuperación de eficacia biológica. Se muestran los valores de aLRT de apoyo estadístico para cada nodo. A su vez, se muestran las ramas con mayor apoyo estadístico con un mayor grosor. Se utilizó el virus original s61 (ver Materiales y Métodos) como grupo externo para enraizar el árbol. La barra muestra la escala en mutaciones por sitio.

Por último, en este árbol se observan muy claramente los diferentes caminos que han seguido las réplicas del mismo virus pasadas en paralelo G1 y G1.5. El árbol nos muestra una separación de estas dos réplicas a partir del pase 21, lo que concuerda con el diferente comportamiento que han seguido. A partir de este pase, el virus G1 llegó a ser el mejor virus en el pase 31, mientras que G1.5 cayó de eficacia muy significativamente en este pase después de ser de los mejores en el pase 21.

Cada virus sigue un camino evolutivo diferente y su progreso concuerda con las características exhibidas. Así, virus con evolución constante recuperan eficacia de forma constante (E1.5 o H1.5), mientras que virus que evolucionan por diferentes caminos tienen recuperaciones más abruptas (K1 o virus del linaje I). Finalmente, el virus K2 ha evolucionado bastante en pase 21 y 31, pero parece no haber encontrado aún el camino de la recuperación. El análisis filogenético de los virus estudiados corrobora los datos previamente obtenidos de una evolución muy heterogénea de los diferentes virus estudiados.

4.3.3 Cálculo de la distancia de cada virus a la raíz del árbol estimado por Máxima Verosimilitud.

Utilizando el árbol estimado y con la ayuda del programa TreeStat, se pudo calcular la distancia de cada virus a la raíz del árbol, generada con el virus original s61 (Figura 4.12). Esta medida nos permite cuantificar la acumulación de mutaciones en los distintos virus, teniendo en cuenta su evolución filogenética. Como se puede comprobar en la gráfica de la Figura 4.13A, la distancia a la raíz de los pases aumentó significativamente en los 10 últimos pases, siendo los virus del pase 31 los que presentan siempre las mayores distancias a la raíz. Además, al comprobar la relación entre las distancias calculadas y los valores de eficacia, se obtuvo una correlación estadísticamente significativa, utilizando el método de Pearson (Figura 4.13B).

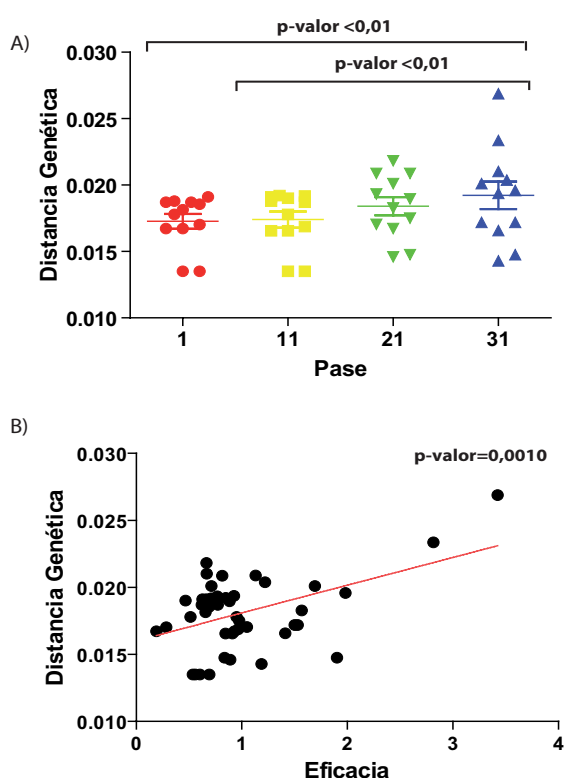


Figura 4.13 Distancia a la raíz del árbol obtenido por Máxima Verosimilitud de los virus en los diferentes pases en gran tamaño poblacional. A) Valores de distancia genética, medida como número de sustituciones por sitio, a la raíz del árbol obtenido por el método de Máxima Verosimilitud. La raíz del árbol filogenético se obtuvo utilizando el virus s61 original como grupo externo. Se muestran los p-valores significativos de la comparación dos a dos de cada pase utilizando la corrección de Bonferroni. Los diferentes pases se representaron siguiendo el código de la Figura 4.1. B) Análisis de correlación entre la distancia a la raíz de cada virus y los valores de eficacia obtenidos. Se muestra el p-valor al utilizar la correlación de Pearson para medir la relación entre las dos variables.

Estos resultados corroboran los saltos evolutivos importantes producidos en los virus estudiados durante los pases de recuperación de eficacia, en concordancia con el aumento de las mutaciones fijadas a partir del pase 21. Esta acumulación de mutaciones se vio asociada al aumento de eficacia biológica.

4.3.4 Cálculo de las tasas de evolución de los virus estudiados y su relación con la recuperación de eficacia.

El siguiente paso en el análisis filogenético del proceso de recuperación de eficacia fue el cálculo de las tasas de evolución de todos los virus en cada pase y la relación de este parámetro con los valores de eficacia biológica observados.

Para llevar a cabo el citado cálculo, se utilizaron métodos de estimación bayesianos, para obtener valores más cercanos a la realidad teniendo en cuenta el modelo evolutivo seguido por los virus, pero sin que estas constricciones condicionasen en exceso los resultados. Se empleó el programa Beast y sus programas accesorios para llevar a cabo las mencionadas estimaciones. En dicho programa, se tuvo en cuenta el pase de cada una de las secuencias, se asignó el modelo GTR+I+G, se adoptó un reloj molecular relajado con una distribución logNormal y se asumió una población constante. Se iteraron dos cadenas de Markov independientes de 20 millones de generaciones cada una, llegando ambas a valores de tamaño de población efectiva mayores de 100 en todos los parámetros, combinándose después para obtener un fichero con 40.000 árboles en total, de los que se eliminaron los 4.000 primeros para generar el árbol consenso sólo con los árboles obtenidos en la fase estable de la búsqueda (Figura 4.14). Una vez obtenido dicho árbol, se representó la tasa de evolución de cada una de las ramas y se procedió a calcular la tasa de cada uno de los virus. Para ello, se estimó el número de sustituciones por sitio y por pase de las ramas externas (correspondientes a los virus estudiados) utilizando el programa TreeStat, obteniéndose de esta manera las tasas de evolución por pase de cada virus en los pases 1, 11, 21 y 31.

Con las tasas de evolución de cada virus por pase calculadas, se analizó estadísticamente la variación de este parámetro a lo largo de los pases. Utilizando una prueba ANOVA de medidas repetidas con una comparación dos a dos corregida por Bonferroni, no se obtuvo un aumento significativo de la tasa de evolución durante los pases de recuperación (Figura 4.15A). Este resultado era debido a la gran dispersión de los valores observada en los pases 21 y 31. Sin embargo los valores medios habían subido y si se comparaban únicamente los valores iniciales y finales con una prueba no paramétrica de muestras repetidas de Wilcoxon, se comprobaba que este aumento era significativo.

Sin embargo, el resultado más importante se obtuvo cuando se analizó la relación entre los valores de la tasa de evolución por pase y los de eficacia durante los pases en gran tamaño poblacional. Para este efecto, se probó la correlación entre dichos parámetros. Utilizando la prueba de Pearson se obtuvo una correlación muy significativa entre la tasa de evolución y la eficacia biológica con un $p\text{-valor} < 0,0001$ (Figura 4.15B). Este resultado indicaba que cuantos más cambios por pase acumulaban los virus, mayor era su eficacia.

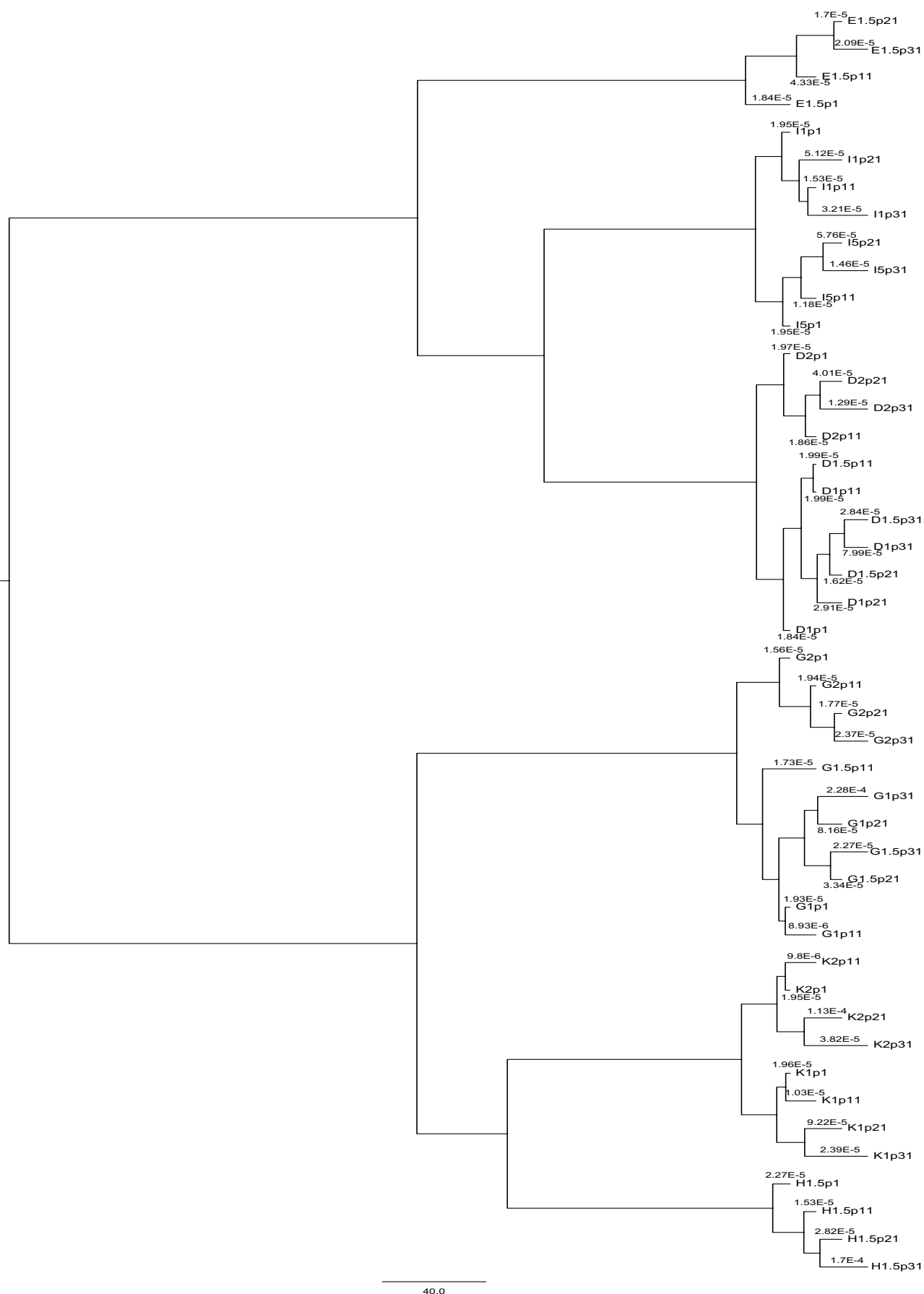


Figura 4.14 Árbol filogenético obtenido por métodos bayesianos. Se utilizaron todos los virus en los diferentes pases de recuperación de eficacia biológica. Se asumió un reloj molecular relajado y se muestran los valores de tasa de evolución (número de sustituciones por sitio y por pase) de las ramas externas, correspondientes a los virus estudiados. La escala representada corresponde a la distancia genética en número de sustituciones por sitio.

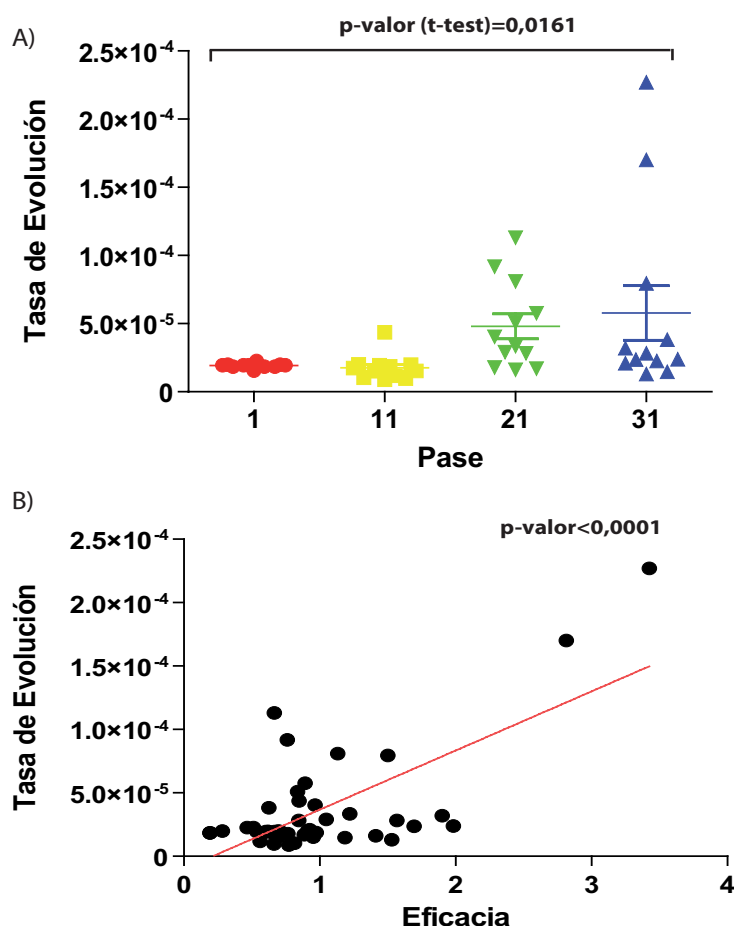


Figura 4.15 Tasas de evolución estimadas para todos los virus en los diferentes pases de recuperación. A) Valores de tasa de evolución de cada virus, medida como número de sustituciones por sitio y por pase, extraídos del árbol filogenético obtenido por métodos bayesianos. Los distintos pases se representan siguiendo el código de la Figura 4.1. Se muestra el p-valor de la comparación entre el pase 1 y el 31 utilizando una prueba t de Wilcoxon. B) Correlación entre la tasa de evolución y los valores de eficacia biológica. Se muestra el p-valor de la prueba de correlación de Pearson.

4.3.5 Análisis por redes de la evolución de las cuasiespecies virales.

Tras realizar los análisis filogenéticos utilizando las secuencias globales completas de los virus, se llevaron a cabo análisis de las cuasiespecies virales de cada uno de ellos. El objetivo de estos análisis era conocer con mayor detalle la evolución de las cuasiespecies a lo largo de la recuperación de eficacia.

Para llevar a cabo estos análisis, se realizó un abordaje diferente y en lugar de utilizar métodos filogenéticos de árboles, se aplicaron métodos de redes. Estos métodos permiten conocer todas las posibles relaciones evolutivas entre las secuencias, en vez de las más probables como ocurre en los árboles, y así lograr una mejor visualización de la evolución de las cuasiespecies. Sin embargo, estos métodos sólo se pueden aplicar dentro de las cuasiespecies relacionadas entre sí, ya que si se aplican a secuencias más distantes, el resultado obtenido es demasiado complejo, generándose redes muy complicadas o hipercubos de caminos imposibles de analizar.

Por todas las razones anteriormente esgrimidas, se decidió utilizar diferentes métodos filogenéticos de redes para analizar la evolución de las cuasiespecies a lo largo de los pases de recuperación. Para ello, se utilizaron las secuencias del espectro de mutantes en las regiones anteriormente analizadas. Se realizaron los análisis por virus o por linajes para no generar demasiada complejidad. Los principales métodos de redes utilizados fueron: el de red de medianas, que genera nodos con los caminos posibles entre las secuencias, utilizando el número de diferencias o distancia de Hamming, y redes con parsimonia, que al utilizar el método de parsimonia, tienen en cuenta las delecciones para generar la red de las secuencias analizadas.

Los diferentes virus en las regiones estudiadas presentaban básicamente dos patrones:

- El primer patrón era el de generación de variabilidad a lo largo de los pases sin desplazamiento de la población hacia otro punto del paisaje de secuencias. Como ejemplo se puede ver la Figura 4.16A, donde se muestran los clones obtenidos durante los pases de recuperación del virus I1 en la región de la LTR-*gag*, en el que a medida que se realizaban los pases se generaba un espectro de mutantes cada vez mayor, pero siempre originándose desde el mismo nodo. Se estaba generando complejidad en la población viral, pero no surgía ningún variante que se impusiese al resto.
- El segundo patrón observado era el de generación de variabilidad, seguido de un desplazamiento de la población hacia otro punto del paisaje de secuencias, que podía ser parcial, como en la región LTR-*gag* del virus I5 (Figura 4.16B), o total, como ocurre en la misma región de los virus H1.5 o G1 (Figura 4.16C y D), con un posterior aumento de variabilidad. En este caso la generación de complejidad seguía siendo predominante, pero surgía algún mutante que era capaz de desplazar a la población, probablemente por su efecto beneficioso en la eficacia biológica. Una vez encontrado este camino, el virus seguía aumentando la nube de mutantes alrededor de la nueva consenso. Este patrón fue detectado en todos los virus en al menos una región, por lo que parecía el mecanismo habitual por el que se estaban fijando las mutaciones a lo largo de los pases.

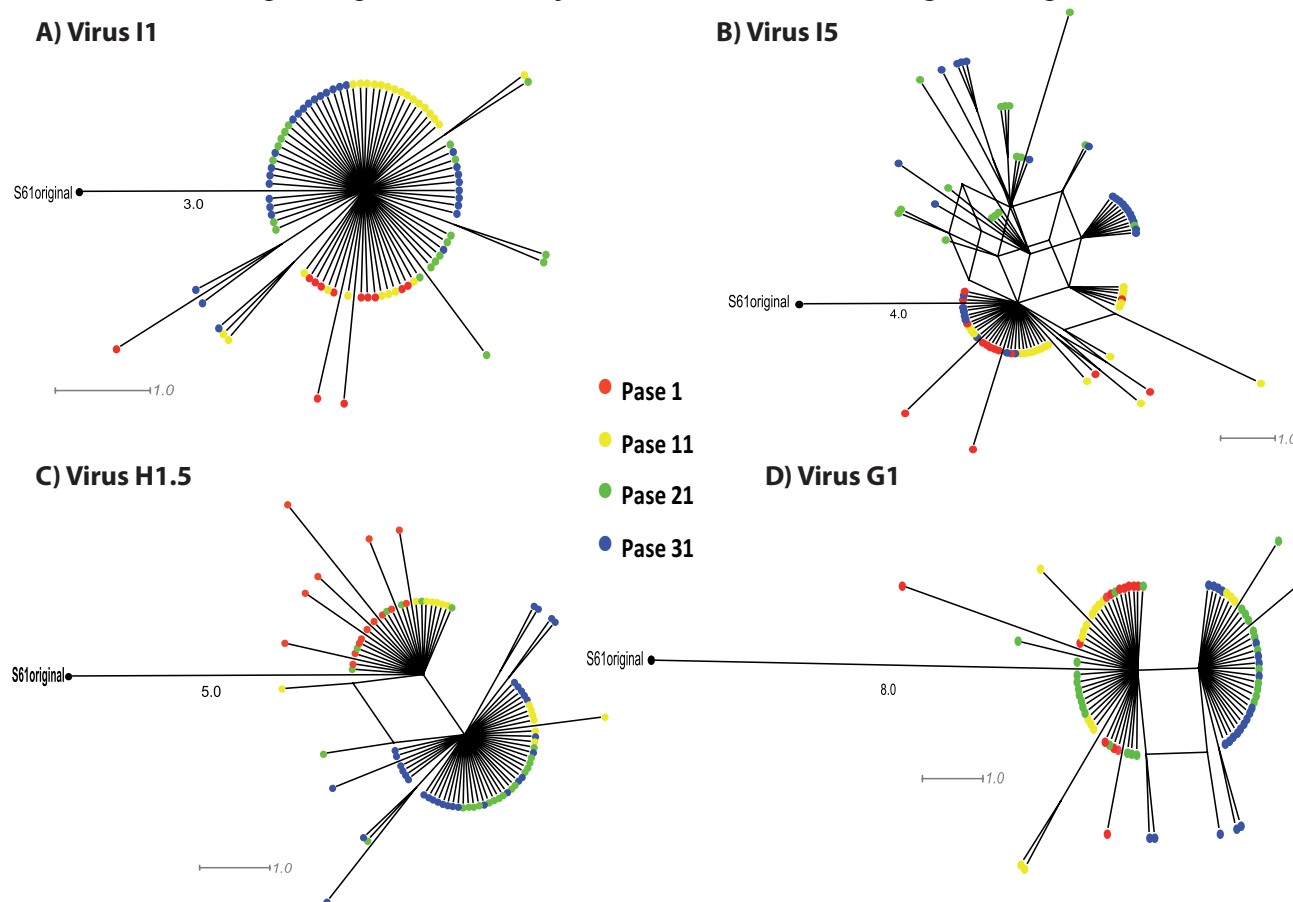
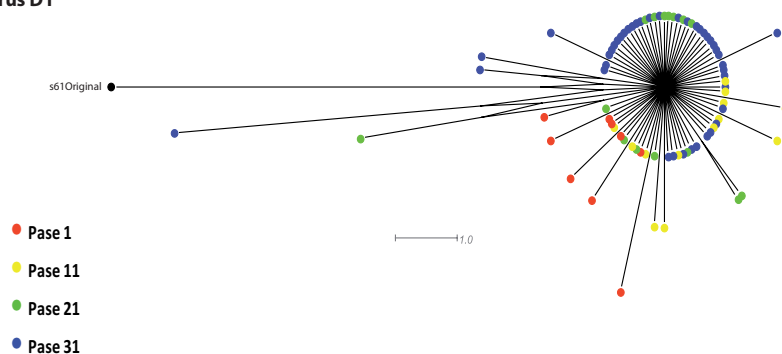


Figura 4.16 Análisis de redes de las cuasiespecies virales en la región LTR-*gag*. Se muestran cuatro ejemplos representativos de los análisis de redes llevados a cabo en las cuasiespecies de los diferentes virus. El método utilizado fue el de red de medianas. Las redes representan el número de diferencias entre las secuencias y la escala corresponde a la longitud de una sola mutación de diferencia. Los nodos se diferenciaron según la secuencia más reciente que contenían, como se indica en la leyenda, y pueden contener una o varias secuencias. También se muestra el número de diferencias con respecto al nodo del virus s61 original. Para generar estas redes se utilizó el programa SplitsTree4.

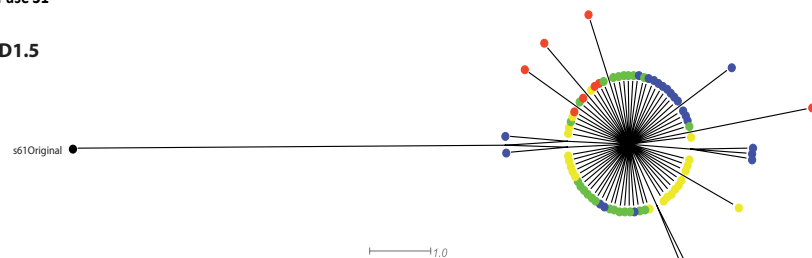
Por otro lado, se utilizaron estos análisis para tratar de entender las diferencias de comportamiento observadas entre las réplicas de los virus pasadas en paralelo D1-D1.5 y G1-G1.5 y entre los virus del mismo linaje K1 y K2. Como recordatorio, en los primeros dos casos, los virus pasados en 5 millones de células MT-4 perdieron eficacia en el pase 31, mientras que en el tercer caso el virus K1 había sido capaz de recuperar su eficacia a lo largo de los pases y el K2 no.

En la comparación entre los dos paralelos arriba mencionados, se observó en las redes obtenidas una variabilidad mucho menor en los virus que perdían eficacia en el pase 31, D1.5 (Figura 4.17) y G1.5 (Figura 4.18), comparados con sus respectivos paralelos D1 y G1. Esto se pudo observar en las diferentes regiones estudiadas, presentándose en las Figuras 4.17 y 4.18 los resultados obtenidos en la región V1-V2 de *env* por ser en los que mejor se visualiza. Estos resultados estaban corroborando la pérdida de heterogeneidad observada en estos virus.

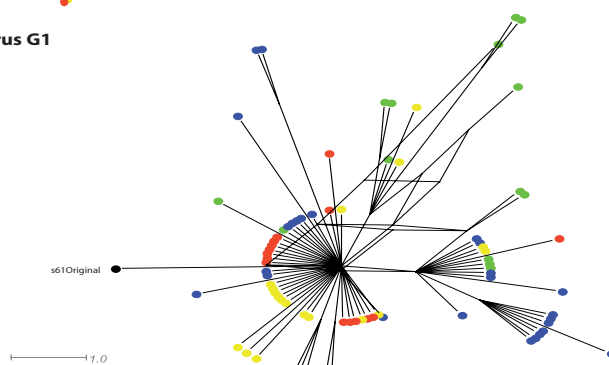
Virus D1



Virus D1.5



Virus G1



Virus G1.5

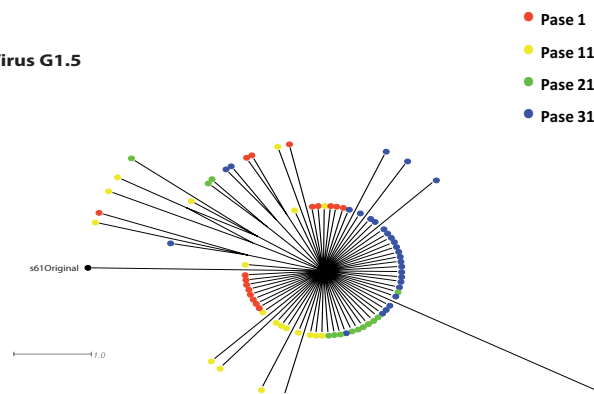


Figura 4.17 Comparación de las redes obtenidas con el virus D1 y D1.5 en la región V1-V2 de la envuelta por el método de red de medianas. Las distancias están medidas en número de diferencias entre nodos. Cada nodo se clasifica según las secuencias más recientes que presenten, como se muestra en la leyenda.

Figura 4.18 Comparación de las redes obtenidas con el virus G1 y G1.5 en la región V1-V2 de la envuelta por el método de red de medianas. Las distancias están medidas en número de diferencias entre nodos. Cada nodo se clasifica según las secuencias más recientes que presenten, como se muestra en la leyenda.

Por su parte, el análisis que se pudo realizar con los dos virus del linaje K fue más exhaustivo. Este linaje había acumulado varias mutaciones en el gen *gag* durante el Trinquete de Muller, entre las que estaba la R15K de la p17. Sin embargo, los dos virus se habían diferenciado entre ellos a lo largo de los pases en gran tamaño poblacional en dos mutaciones también en la p17, la G12R y la Q63H. Estos dos cambios los presentaba el virus K1, que recuperaba eficacia, mientras que el virus K2, que había sido incapaz de recuperarse, no los había adquirido. Al analizar las redes generadas en la región de la p17, se podía observar cómo la población de K1 se desplazaba completamente, debido a genomas que tenían las dos mutaciones fijadas (Figura 4.19). Esta imposición de los dobles mutantes, coincidía con la fuerte subida de eficacia que el virus K1 experimentaba en los últimos pases. Además, la frecuencia con que aparecían las dos mutaciones juntas dentro de la cuasiespecie describía una curva exponencial, indicando un aumento de la frecuencia de los dobles mutantes, una vez que se había acumulado una mutación sobre la otra. Por otro lado, al observar la red estimada para el virus K2, se podía comprobar cómo aún no se había fijado la doble mutación en la población y sólo había clones con una de las dos mutaciones, excepto un único caso en el que ya aparecían las dos. Sería de esperar que en futuros pases la descendencia de este clon se impusiese en la población.

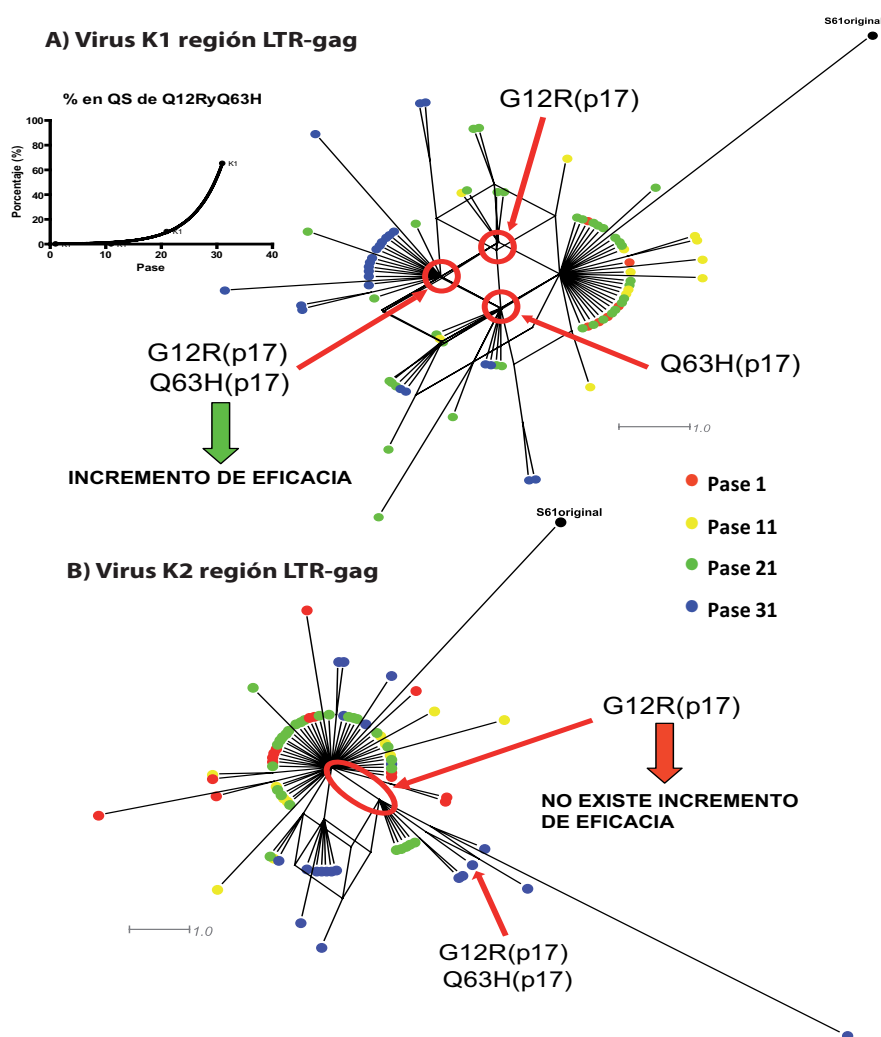


Figura 4.19 Comparación de las redes obtenidas por el método de red de medianas con los virus K1 y K2 en la región LTR-*gag* a lo largo de los pases en gran tamaño poblacional. A) Cuasiespecies del virus K1 en las que se muestra la imposición del doble mutante en p17 G12R-Q63H en la población y su asociación con el incremento de eficacia observado en los últimos 10 pases. La gráfica representada corresponde a la fijación exponencial en las cuasiespecies de variantes con la doble mutación. Se representa el porcentaje de dobles mutantes presentes en la cuasiespecie a lo largo de los pases de recuperación. B) Cuasiespecies del virus K2 en las que se indica que sólo aparecen mutantes sencillos G12R, no habiéndose llegado a imponer en la población. Únicamente se ha observado un clon con la doble mutación detectada en K1 (señalado como G12R-Q63H). En este virus no se ha podido observar un aumento de eficacia biológica. En ambos casos, cada nodo se clasifica según las secuencias más recientes que presenten, como se muestra en la leyenda.

Según los resultados obtenidos, lo más probable es que el cambio R15K, aparecido en la proteína p17 durante los pases placa a placa, sea perjudicial para la eficacia del virus y se compense, a lo largo de los pases en gran tamaño poblacional, con el cambio G12R. Este cambio no tendría la eficacia suficiente para imponerse en la población y necesitaría complementarse con el Q63H, que ya permitiría a estos variantes imponerse en la población y recuperar eficacia.

4.4 Caracterización fenotípica.

Una vez analizadas las características genéticas asociadas con la recuperación de eficacia biológica, se decidió estudiar las características fenotípicas de los virus estudiados y su evolución a lo largo del proceso.

4.4.1 Estudio de la citopatogenicidad de los virus a lo largo de la recuperación de la eficacia biológica.

En primer lugar se realizó un ensayo para determinar el efecto citopático de los virus. Para ello, se infectaron en un mismo experimento, 10^5 células MT-4 con los virus recogidos en el pase 1, 11, 21 y 31 a una m.o.i. de 0,01, realizándose fotos de la evolución de los cultivos y mediciones de las viabilidades celulares. De esta forma, se pudo observar en los virus estudiados un clarísimo aumento de la citopatogenicidad a lo largo de los pases. Las fotos de los cultivos celulares mostraban una aparición de sincitios más temprana en los virus de pases más avanzados. Además, se visualizaron mayores efectos citopáticos como un mayor número de sincitios, sincitios mucho más grandes o mayor disgregación celular (Figura 4.20). Por otra parte, se podían distinguir virus con un efecto más sincitial (por ejemplo el linaje D, Figura 4.21) o con efecto más apoptótico (linaje G, Figura 4.21). Pero, en general, se observó una mayor citopatogenicidad en todos los virus en pases más avanzados.

Asimismo, las medidas de viabilidad de los cultivos celulares apoyaron este resultado. Al tercer día post-infección, se obtuvieron unos valores de viabilidad celular significativamente menores en el pase 21 y, sobre todo, en el pase 31 respecto a los pases anteriores (Figura 4.22). En ese momento, todos los virus del pase 31 presentaban valores inferiores al 40% de células vivas. Estos resultados indicaban un aumento muy significativo de la citopatogenicidad de los virus a medida que eran pasados en grandes tamaños poblacionales, produciendo la muerte celular de los cultivos cada vez con mayor rapidez.

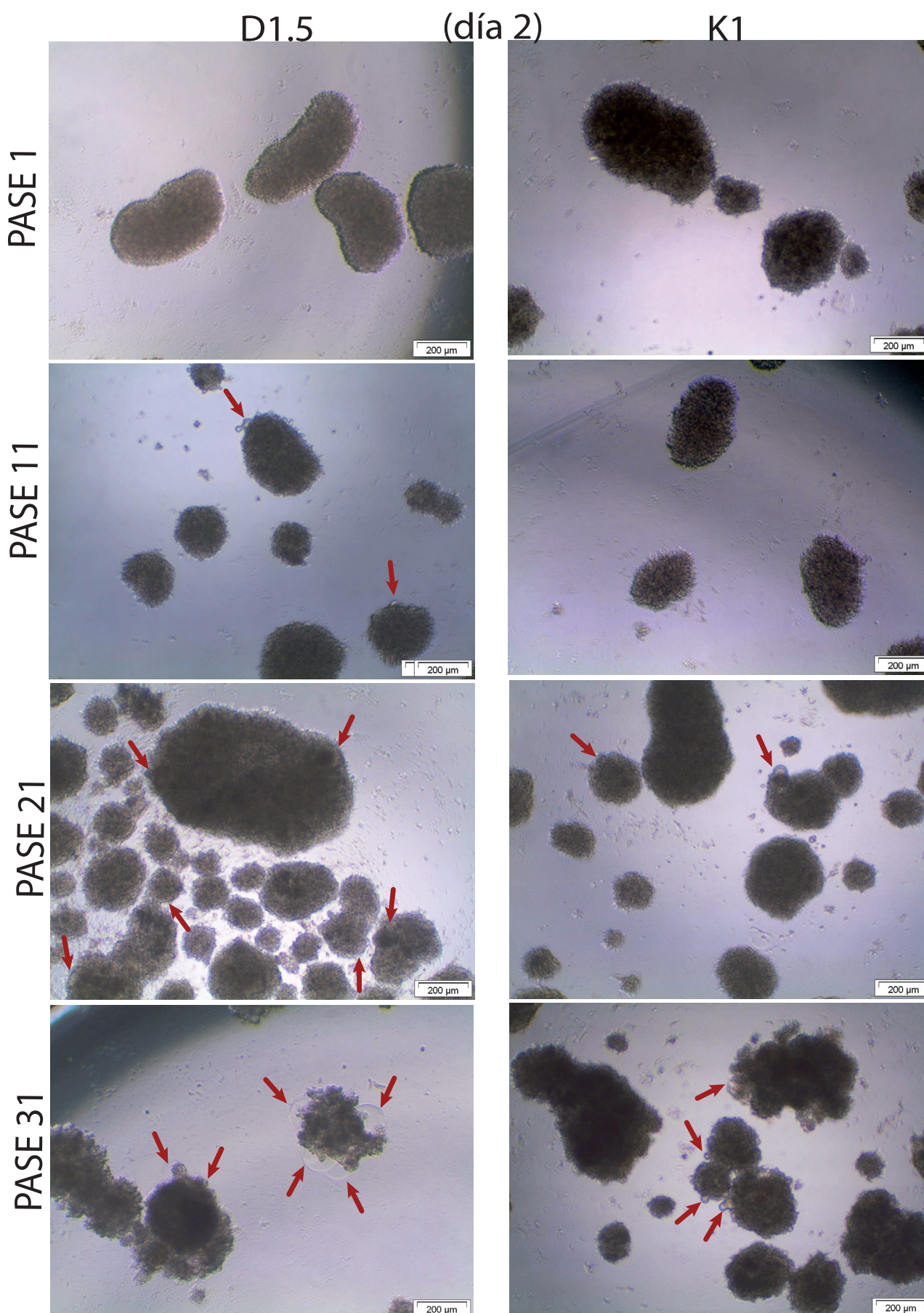
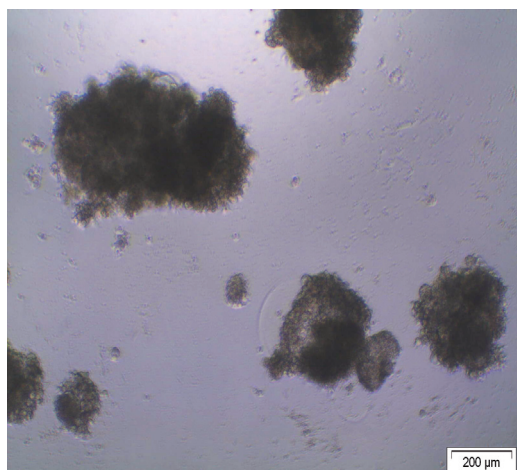
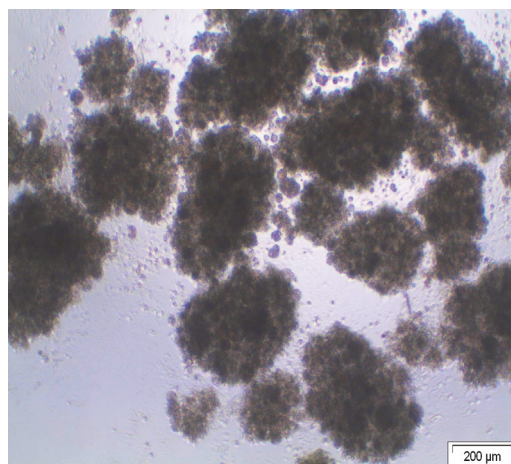


Figura 4.20 Ensayo de citopatogenicidad con todos los virus en los diferentes pases. Se muestran, a modo de ejemplo, fotos de los cultivos infectados, a día 2 post-infección, de dos de los virus estudiados, D1.5 y K1, obtenidos en el pase 1, 11, 21 y 31. Se puede observar el aumento de la citopatogenicidad a medida que aumentan los pases. Las flechas muestran algunos de los sincitios observados.

Diferentes Fenotipos



D1p31 (Sincitial)



G1p31 (Apoptótico)

Figura 4.21 Diferencia entre fenotipos observados en virus del pase 31 de recuperación de eficacia a día 2 post-infección. Se muestra un ejemplo de infección con virus de fenotipo más sincitial (D1p31), con grandes sincitios celulares, y un ejemplo de un virus con fenotipo más apoptótico (G1p31), con gran cantidad de células muertas por apoptosis.

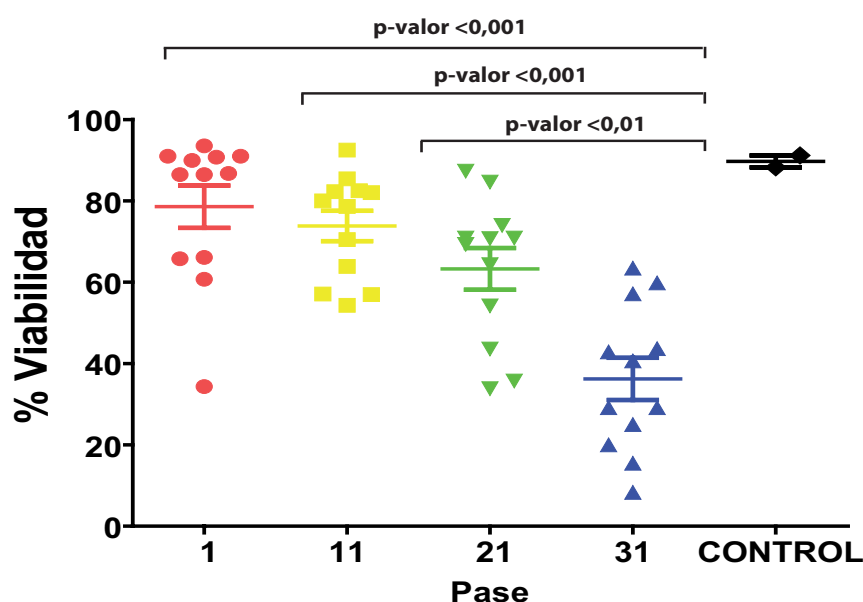


Figura 4.22 Porcentaje de viabilidad celular, 3 días post-infección, en cultivos de células MT-4 infectados con todos los virus en los diferentes pases de recuperación. Se muestra el valor medio de viabilidad de los cultivos celulares infectados con cada virus. Estas infecciones se llevaron a cabo todas en paralelo para poder compararlas. También se muestra el control negativo que constituyó el cultivo simultáneo de células MT-4 sin infectar. Los distintos pases se representan siguiendo el código de la Figura 4.1. Se muestran los p-valores significativos de la comparación dos a dos de cada pase utilizando la corrección de Bonferroni.

4.4.2 Determinación del título viral y de la producción de proteína p24.

El siguiente paso para caracterizar propiedades fenotípicas de los virus estudiados fue analizar tanto los títulos virales como la producción de proteína p24. El primer dato nos permitía conocer la infectividad de los virus, mientras que el segundo nos ofrecía datos de producción de proteínas virales e indirectamente de partículas virales, ya que p24 es una de las proteínas que componen la cápside viral.

Los valores de título viral mostraban una ligera tendencia a incrementar las dosis infectivas a lo largo de los pases hasta el pase 21, produciéndose posteriormente una caída de estos valores

en el pase final (Figura 4.23A). Sin embargo, debido a la dispersión de los datos, ninguna de estas diferencias era estadísticamente significativa.

Por su parte, la producción de proteína p24 se ha mantenido constante hasta el pase 21, aunque también con una ligera tendencia a aumentar, y con una optimización cada vez mayor de los valores, como se puede derivar de la disminución de las barras que representan la desviación estándar (Figura 4.23B). No obstante, los valores cayeron significativamente en el pase 31, con una disminución estadísticamente significativa entre el pase 21 y 31.

En resumen, estos resultados indicarían que los virus en su proceso de recuperación han llegado a un óptimo de infectividad y producción viral tras 21 pases. Sin embargo, a pase 31 la producción viral se ha visto disminuida probablemente asociada con la menor viabilidad de los cultivos infectados con los virus del pase 31.

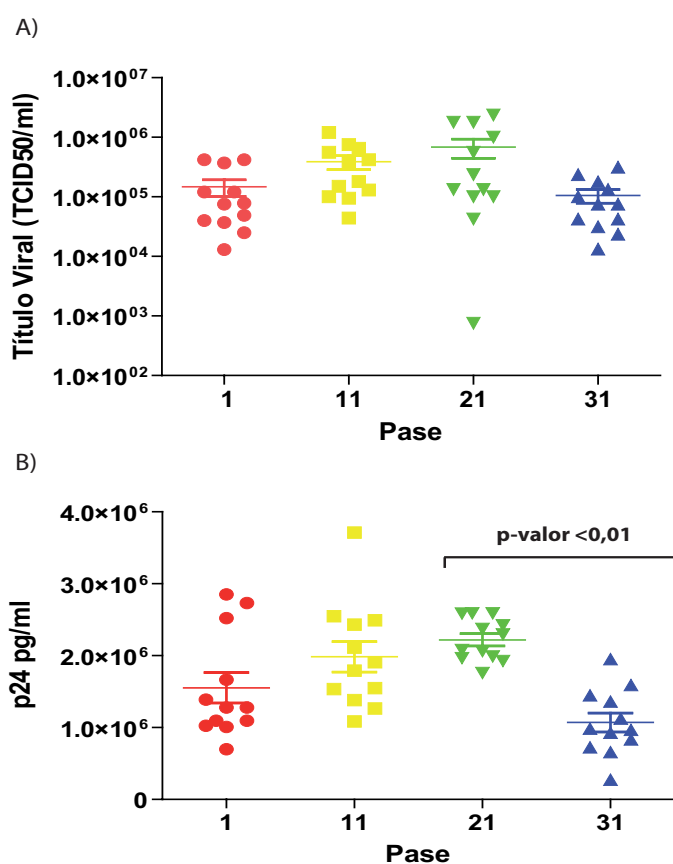


Figura 4.23 Valores de título viral y producción de proteína p24 a lo largo de los pases en gran tamaño poblacional. A) Valores de título viral en los pases 1, 11, 21 y 31. B) Valores de producción de la proteína de la cápside viral p24 durante los pases de recuperación. En ambas figuras los distintos pases se representan como en 4.1. Se muestra el p-valor significativo en las comparaciones dos a dos de los distintos pases utilizando la corrección de Bonferroni.

4.4.3 Análisis la actividad RT a lo largo de los pases de recuperación.

Otra característica fenotípica que se estudió fue la actividad de la RT de los sobrenadantes de los cultivos infectados por virus en los diferentes pases. Esto se hizo mediante un ensayo de RT (ver Materiales y Métodos apartado 4.6.4). El experimento se realizó con todos los virus de los diferentes pases en un mismo ensayo, para una mejor comparación de los valores.

La actividad RT de los virus ha ido decayendo a lo largo de los pases, hasta llegar a su valor mínimo en el pase 21 (Figura 4.24A), repuntando algo los valores en el pase 31. De igual forma, al normalizar los valores de actividad RT por los de producción de p24 para comprobar la ratio RT por partículas virales, el resultado fue el mismo, una caída continua hasta el pase 21, seguida de un

repunte, en este caso aun mayor, en el pase 31 (Figura 4.24B). Finalmente, los valores de actividad RT divididos por título viral, para estudiar la actividad de las RT de las partículas infecciosas, aportó un resultado distinto, no cambiando los valores medios a lo largo de los pases (Figura 4.24C).

De forma inesperada, la actividad RT total de los virus ha ido disminuyendo a lo largo de los pases de recuperación con un repunte en los 10 últimos pases. Sin embargo, la actividad RT de las partículas infecciosas se ha mantenido invariable.

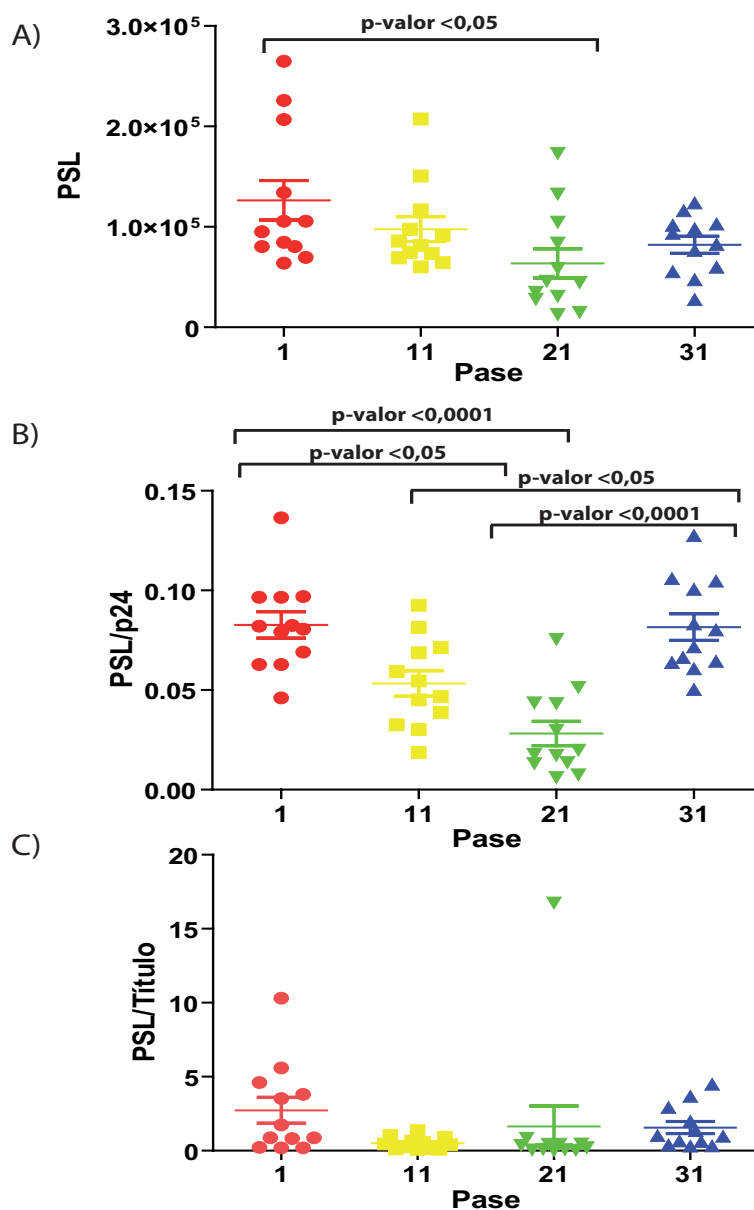


Figura 4.24 Actividad retrotranscriptasa de los virus durante los pases de recuperación. A) Valores de actividad retrotranscriptasa medidos en unidades de luminiscencia foto-estimulada (PSL) de ensayos llevados en paralelo con todos los virus a la vez. Los valores totales se obtuvieron al revelar una placa fotográfica expuesta a la membrana en la que se realizó el ensayo de actividad retrotranscriptasa con ^{32}P . B) Valores de actividad retrotranscriptasa normalizados por producción de proteína p24 y C) por título viral. Los diferentes pases se representan como en 4.1. Se muestran los p-valores estadísticamente significativos en las comparaciones dos a dos de los distintos pases utilizando la corrección de Bonferroni.

4.5 Análisis multivariable por regresión lineal de la recuperación de la eficacia biológica.

Una vez completados todos los análisis sobre la recuperación de eficacia biológica anteriormente descritos, se disponía de una completa caracterización de los virus estudiados con numerosas variables analizadas. Debido a este amplio conocimiento de las características de los virus, se estaba en disposición de intentar modelizar el proceso de recuperación de eficacia realizando un análisis por regresión utilizando todas las variables de los que se disponía. Con este estudio se pretendía conocer cuáles de las mencionadas variables estaban directamente relacionadas con la eficacia biológica y su peso en la variación de la misma.

Con este objetivo, se realizaron los análisis de regresión, utilizando como variable dependiente la eficacia y como variables independientes las siguientes:

- **Número de mutaciones.**
- **Porcentaje del virus completo en V1-V2.**
- **Heterogeneidad de las cuasiespecies.**
- **dN-dS calculado en las cuasiespecies.**
- **Título viral.**
- **Producción de proteína p24.**
- **Actividad RT dividida por p24.**
- **Distancia genética a la raíz.**
- **Tasa de Evolución de las ramas de cada virus.**

El **número de mutaciones** enumeraba los cambios presentes en las secuencias globales con respecto a los virus iniciales, a los que se asignó el valor 0. Con el **porcentaje del virus completo en V1-V2** presente en las poblaciones, se pretendía probar la importancia de la acumulación de deleciones en esta región. Se calculó el porcentaje de completos en vez del de las deleciones, ya que cada virus acumulaba diferentes variantes, lo que hacía su representación muy difícil. La **heterogeneidad de las cuasiespecies** reflejaba la complejidad del espectro de mutantes de cada virus. Esta variable fue calculada como media de la distancia genética entre los variantes de la cuasiespecie en número de sustituciones por sitio. La diferencia **dN-dS calculada en las cuasiespecies** fue utilizada para comprobar la existencia de selección positiva dentro de la nube de mutantes y su efecto sobre la eficacia. Se incluyó el **título viral** como medida de la infectividad de los virus. La **producción de proteína p24** daba una idea de la producción de partículas virales. Se utilizó la **actividad RT dividida por p24** para tener esta medida normalizada por la cantidad de partículas virales presentes, lo que daría una idea más aproximada de la calidad media de la RT de los virus estudiados. La **distancia genética a la raíz** era el cálculo de la distancia genética, medida en número de sustituciones por sitio y estimada por Máxima Verosimilitud, a la raíz generada por el virus s61 en el árbol estimado para la recuperación de eficacia. Finalmente, la última variable añadida fue la **tasa de evolución de las ramas de cada virus**, medida como número de sustituciones por sitio y por pase.

Para el estudio se asumió un modelo lineal, ya que era el modelo más sencillo y la eficacia parecía aumentar de esta forma a lo largo de los pases. En primer lugar, se incluyeron en el modelo todas las variables y una constante. En este primer paso se quería comprobar qué variables influían significativamente en la variación de la eficacia para eliminar el resto. En esta primera aproximación, las únicas variables con influencia estadísticamente significativa al 90% eran la heterogeneidad, dN-dS y, sobre todo, la tasa de evolución. Esta regresión era significativa con un p -valor = $6,5 \times 10^{-5}$ y un cuadrado del coeficiente de regresión (R^2) de 0,6. Sin embargo en este modelo, el término independiente no era significativo con p -valor de 0,21, lo que permitía asumir una regresión que pasase por el origen.

Al volver a realizar la regresión, en este caso sin término independiente, la regresión mejoró de manera significativa, ya que la desviación estándar de la estimación bajó su valor de 0,44 a 0,42 y la significación del modelo subió hasta un p -valor = $5,8 \times 10^{-15}$. En este modelo, las variables que tenían una influencia estadísticamente significativa, seguían siendo heterogeneidad y la tasa de evolución, ahora ambas con gran significación.

Finalmente, se eliminaron del modelo todas aquellas variables cuya influencia no era significativa, ya que no aportaban casi nada al modelo y lo hacían más complicado. Tras varias rondas de análisis, se decidió mantener únicamente la heterogeneidad y la tasa de evolución en el modelo, eliminándose el resto de las variables. Tanto la heterogeneidad como la tasa de evolución en el modelo final presentaban una significancia mayor del 99,99%, aportando cada variable la mitad del peso en el modelo, con coeficientes estándar de 0,57 para la heterogeneidad y 0,43 para la tasa de evolución. El modelo que incluía la heterogeneidad y la tasa de evolución tenía una R^2 de 0,864, lo que suponía que el 86% de la variación de la eficacia se explicaba con la regresión obtenida, y un error estándar de 0,44 en la estima de la eficacia. Varios análisis posteriores (Figura 4.25) indicaban que el modelo se comportaba razonablemente bien, ya que las frecuencias de los residuos se distribuían normalmente, las probabilidades esperadas y observadas se ajustaban a una recta y ningún punto se alejaba excesivamente del comportamiento general. La recta que definía el modelo final ajustado era:

$$W = 164 \times H + 9 \times 10^4 \times TE$$

p-valor (0,000) (0,000)

Siendo W la eficacia biológica, H la heterogeneidad de las cuasiespecies y TE la tasa de evolución medida en la ramas de cada virus.

El modelo obtenido nos indicaba que las dos variables relacionados con la generación de complejidad poblacional incluidos en el análisis, la heterogeneidad de las cuasiespecies y la tasa de evolución de los virus, eran los que estaban dirigiendo la recuperación de eficacia en el sistema “*in vitro*” estudiado. Este resultado subraya el importante papel que desempeña la complejidad de las poblaciones virales en la eficacia de los virus objeto del presente trabajo.

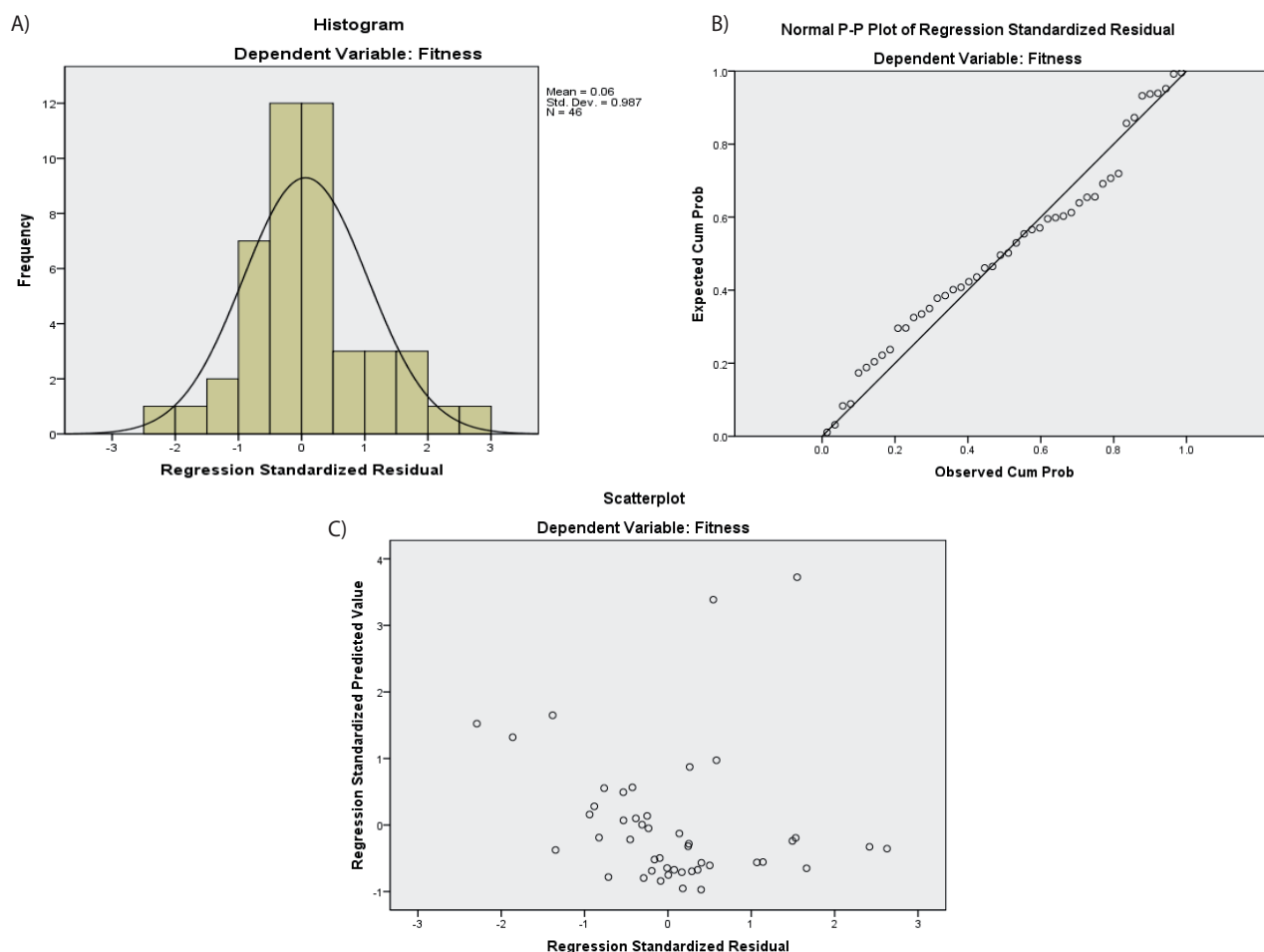


Figura 4.25 Análisis del modelo de regresión múltiple utilizando la heterogeneidad y la tasa de mutación como variables asociadas a la eficacia. La bondad del ajuste del modelo obtenido $W = 164 \times H + 9 \times 104 \times TE$ se probó utilizando diversas representaciones. A) Histograma que representa la frecuencia de los residuos estandarizados del modelo de regresión múltiple utilizado. Conjuntamente con el histograma, se representa la densidad normal asociada a la muestra. Se puede observar cómo la distribución de los residuos se ajusta razonablemente a una distribución normal, acumulándose los valores principalmente en la región central de la curva. B) El diagrama P-P compara la frecuencia acumulada por los residuos tipificados con la probabilidad esperada bajo la hipótesis de normalidad. Aunque algunos valores se alejan algo de la recta, en general se ajustan bastante a la hipótesis de normalidad. C) Diagrama de dispersión de los valores estandarizados predichos frente a los residuos estandarizados. Vemos que no existe ningún patrón entre las dos variables y que ningún punto se aleja excesivamente del resto de los valores, por lo que podemos asumir que el modelo se está comportando de manera aceptable.

4.6 Estudio de mutaciones aparecidas en los experimentos de evolución “in vitro” con diferentes tamaños poblacionales.

4.6.1 Análisis de cambios de aminoácidos aparecidos durante los experimentos de evolución “in vitro”, modelado por homología de sus estructuras y caracterización fenotípica de los virus generados.

Debido a la aparición de mutaciones a lo largo de la pérdida y recuperación de Eficacia Biológica que parecían importantes en el proceso, se decidió estudiar más exhaustivamente e individualmente algunas de ellas. Para ello, se introdujeron en el virus original s61 las mutaciones elegidas, mediante mutagénesis dirigida (Materiales y Métodos, apartado 2.8). Una vez obtenidos los virus con los cambios deseados, se estudiaron sus características comparándolas con las del virus s61 original. De este modo, se analizaron el título, la producción de proteína p24, la citopatogenicidad y la eficacia biológica de cada mutante. Además, se hizo modelado por homología de las correspondientes proteínas (o ARN en el caso de los cambios en LTR) y se compararon con las proteínas originales del s61. Con esta metodología se pudo tener una idea de sus efectos y el papel que jugaban en el proceso de pérdida y recuperación de eficacia biológica.

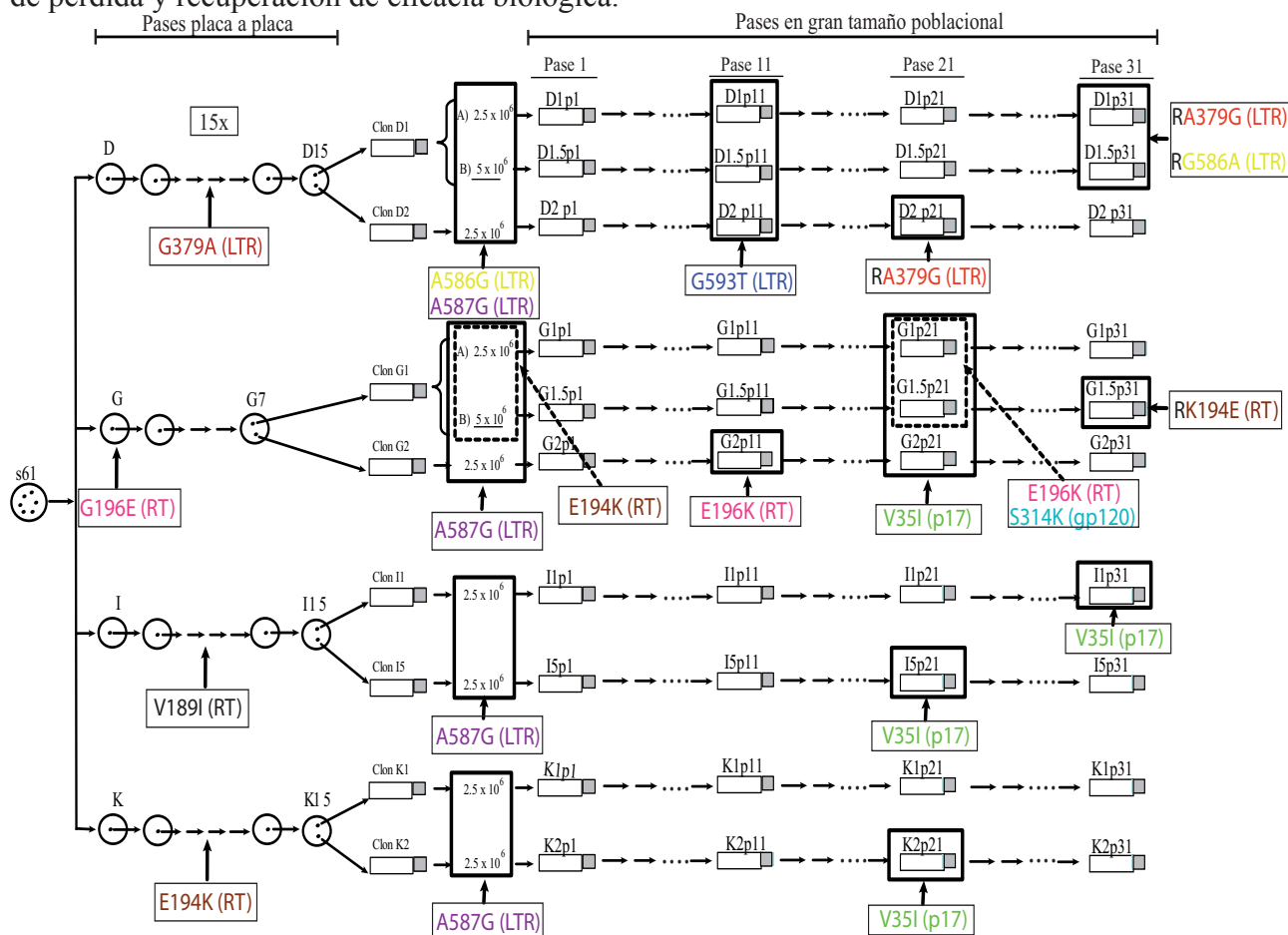


Figura 4.26 Esquema de las mutaciones estudiadas por mutagénesis dirigida. Sobre el esquema general de los pases llevados a cabo con los diferentes clones biológicos, se muestra, con flechas y recuadros las mutaciones seleccionadas para estudiar. Para simplificar en el esquema no se muestran los linajes E y H, ya que no se eligieron ninguna de las mutaciones aparecidas en estos linajes. Se muestran todas las mutaciones, nombradas con el cambio de aminoácido correspondiente, excepto en las de la LTR que se indica el cambio de nucleótido. En cada cambio se señala su momento de aparición y el virus en el que surgieron. Los cambios en la misma posición se muestran con diferentes colores y en el caso de las reversiones, se añade una R en negrita delante. Los cuadrados indican en qué virus se detectaron las mutaciones señaladas. En el caso del linaje G, se muestran tres mutaciones con cuadrados y flechas discontinuas para indicar que estas mutaciones sólo se fijaron en dos de los tres virus del linaje (en todos los casos G1 y G1.5).

Las mutaciones analizadas por mutagénesis dirigida se nombran con el cambio de aminoácido correspondiente, excepto las de LTR que se nombran con el cambio de nucleótido. Dichas mutaciones se dividieron en mutaciones aparecidas en los pases placa a placa, mutaciones aparecidas al hacer la semilla posterior a los cuellos de botella y mutaciones fijadas durante los pases de recuperación de eficacia.

4.6.1.1 Mutaciones aparecidas en los virus durante los pases placa a placa.

G379A (LTR)

Esta mutación aparece en la región U3 de la LTR, en concreto en el sitio de unión del factor de transcripción Sp1-III. Apareció durante los pases seriados en cuello de botella en el linaje D y revirtió en los pases D2p21, D1p31 y D1.5p31 (Figura 4.26). Esta mutación resultaba interesante porque era de las pocas que se habían fijado durante el Trinquete de Muller en el linaje D y se situaba en una zona importante para la replicación viral. En este linaje se fijaron posteriormente varias mutaciones en la LTR en el proceso de recuperación, pudiendo estar condicionadas por esta mutación inicial. Al utilizar un modelado de estructura de ARN con esta mutación y compararla con la original del s61 (Figura 4.27) se pudo observar un cambio fuerte en la estructura, como demuestra el aumento de 9 kcal/mol de su energía libre.

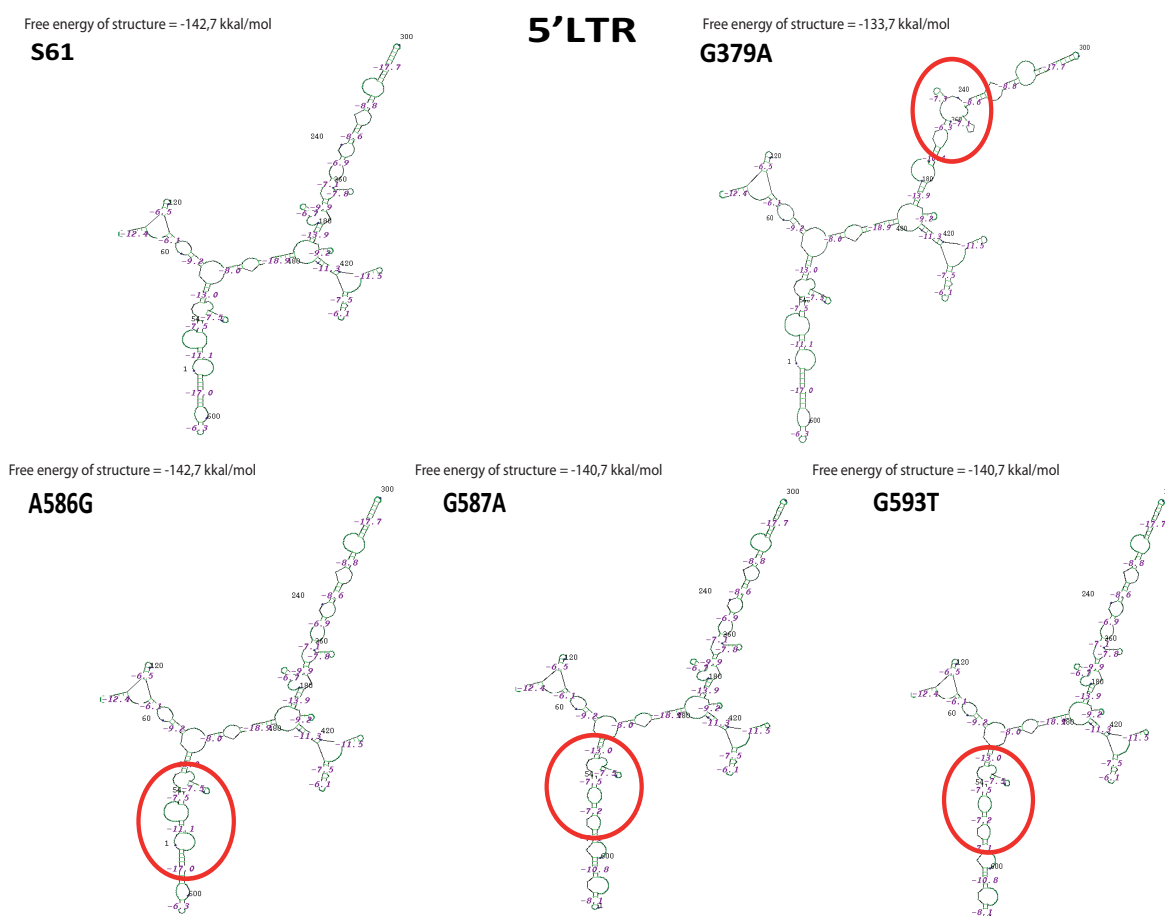


Figura 4.27 Modelado de la estructura secundaria del ARN de la región LTR del virus s61 original y de los mutantes G379A, A586G, G587A y G593T en dicha región. Se muestra la estructura predicha por el programa GeneBee, con la energía libre estimada para cada una de ellas. Los círculos rojos muestran la localización de cada uno de los cambios y la zona adyacente que se ve afectada por dichos cambios.

Esta mutación causaba una caída de eficacia biológica muy significativa con respecto al virus s61 (Figura 4.32). Mientras que la producción de proteína p24 se mantenía similar al control, el título viral presentaba valores más bajos, por lo que esta mutación estaba produciendo mayor cantidad de virus defectivos. Finalmente, la citopatogenicidad de este mutante era similar a la del virus control.

V189I (RT)

Mutación aparecida en la RT del linaje I durante el Trinquete de Muller (Figura 4.26). Este clon sufrió la mayor caída de eficacia en los pases placa a placa, ya que era el virus con mayor valor inicial. Una de las mutaciones que presentaba era ésta, que al estar cerca de la región catalítica de la RT, parecía interesante estudiar. Es una mutación que ha sido asociada a la adquisición de resistencia a la Nevirapina (Balzarini y cols., 1997; Kleim y cols., 1996).

Su modelado por homología mostraba una pequeña alteración estructural, ya que desplazaba ligeramente una hélice α adyacente, denominada F en la estructura de la RT, situada en el lateral de la región de la palma (Figura 4.28). Sin embargo, la zona de interacción con el ADN no presentaba alteraciones.

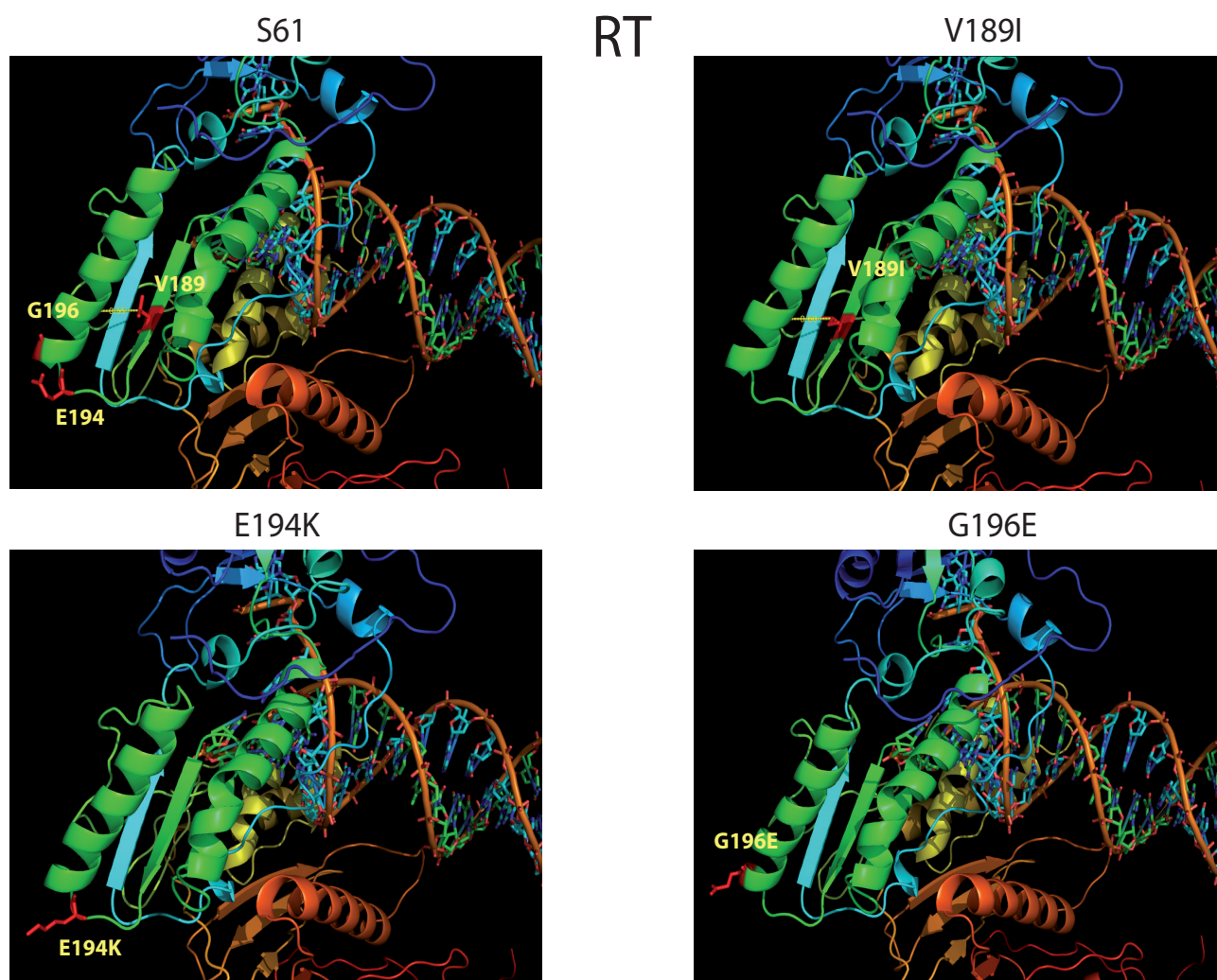


Figura 4.28 Modelado de la RT del virus s61 y de los mutantes V189I, E194K y G196E. En la RT del virus s61 se muestran las posiciones correspondientes a los diferentes mutantes estudiados en esta proteína. En cada mutante se muestra el cambio introducido y el aminoácido de color rojo y representado en forma de barras. En el caso del s61 y del virus V189I, se muestra también, con una línea de unión, la distancia entre la posición 189 y la hélice α adyacente. En el s61 esta distancia era de 8,55 Å y en el V189I de 8,60 Å. En el modelo se incluyó una molécula de ADN para mostrar la parte catalítica y de unión a ADN de la proteína.

La eficacia biológica que presentaba esta mutación era menor que la del s61, aunque no llegaba a ser estadísticamente significativa (Figura 4.32). Su producción de p24 era mucho mayor que la del control, mientras que su título era menor. Su citopatogenicidad era algo menor que la del control.

Esta mutación en la RT suponía una caída en la infectividad viral, ya que la razón título/p24 caía significativamente en este mutante. Por este motivo, se puede considerar esta mutación como perjudicial para el virus.

E194K (RT)

Este cambio aparecía en dos linajes, pero en contextos diferentes. En el linaje K había aparecido durante el Trinquete de Muller y se había mantenido durante los pases de recuperación (Figura 4.26). Estos virus habían tenido una difícil recuperación, de tal modo que uno de ellos, el virus K2, no había sido capaz de recuperarse en los 30 pases. Por su parte, también había aparecido en los virus G1 y G1.5 del linaje G, pero esta vez al hacer la semilla tras los cuellos de botella (Figura 4.26). En este caso, esta mutación aparecía en genomas que presentaban una mutación previamente adquirida, la G196E. Esta mutación revirtió en el virus G1.5 cuando desapareció la E196 con el cambio E196K.

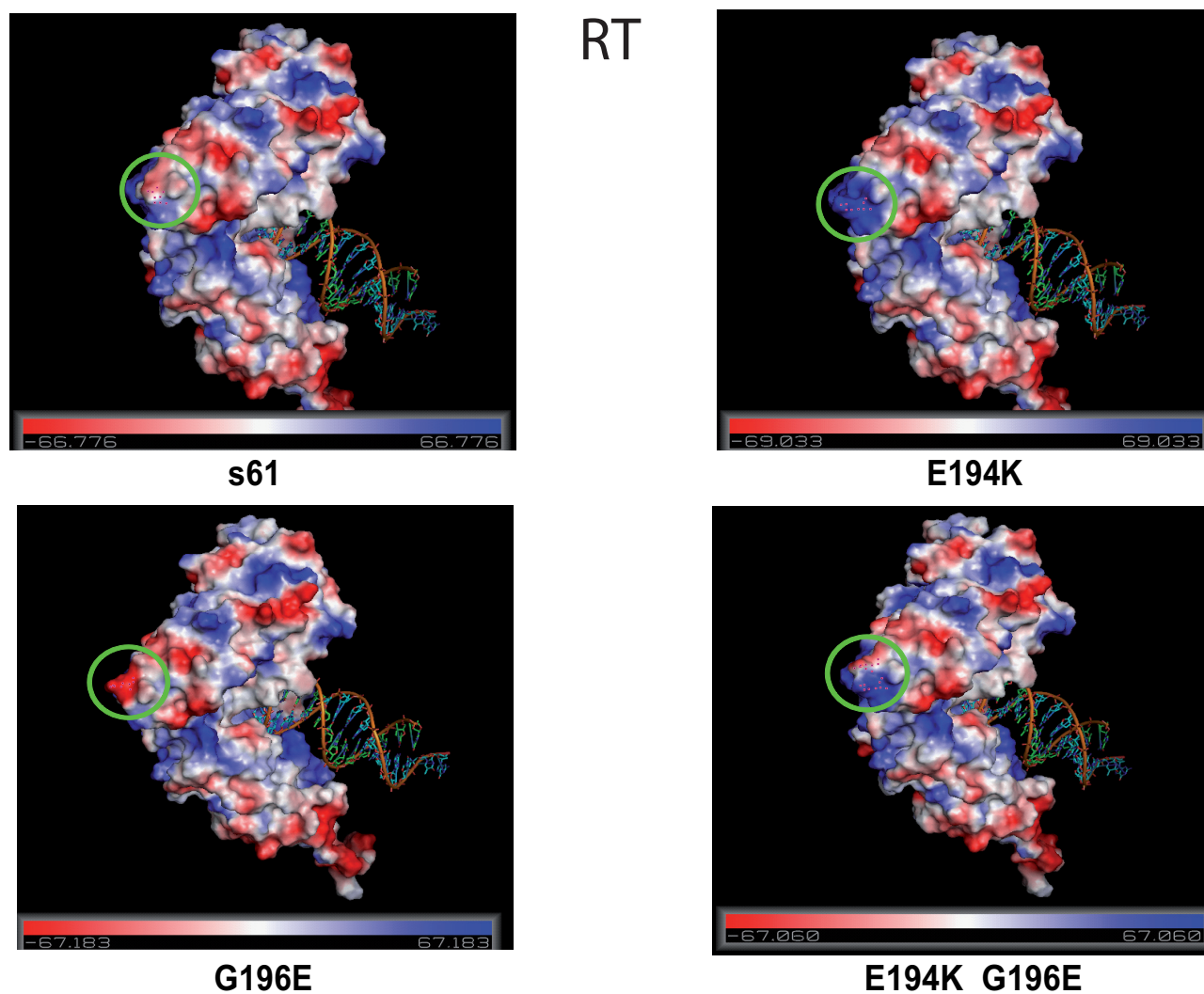


Figura 4.29 Predicción de la carga electrostática de los modelos de la RT del virus s61 y de los mutantes E194K, G196E y E194K-G196E juntos. Se modeló el doble mutante para tener una idea del efecto de las dos mutaciones a la vez. El círculo verde indica la localización de las posiciones 194 y 196 de la proteína. Las regiones apolares se muestran en azul, mientras que las polares aparecen en rojo. Por último, las zonas con carga neutra se muestran en blanco. Se muestra una molécula de ADN unida al heterodímero que conforma la proteína.

Su modelado no mostraba cambios estructurales, aunque el cambio aparecía en la misma hélice α que se desplazaba en la mutación anterior (V189I) (Figura 4.28). Este cambio causaba una alteración considerable de la carga electrostática de la zona, haciéndola más positiva (Figura 4.29).

La eficacia de este mutante era significativamente menor que la del virus s61 (Figura 4.32). Esta caída de eficacia se vio acompañada de un gran aumento de producción de proteína p24 y de una fuerte disminución del título viral. Por su parte, la citopatogenicidad de este virus era mayor que la del control, produciendo la muerte del cultivo celular ya en el día 5 post-infección.

Este mutante mostraba características similares a las del V189I, pero más acentuadas. La caída de eficacia observada se debió a una caída muy fuerte de la infectividad de este virus, lo que indicaría la producción de una enorme cantidad de virus defectivos. Sin embargo, su gran citopatogenicidad podía explicar la fijación de esta mutación en el linaje G, que ya tenía la mutación G196E.

G196E (RT)

Relacionada con la mutación anterior, ésta se encontraba en el clon G original (Figura 4.26). A su alrededor se han ido acumulando mutaciones durante la recuperación de eficacia, como la previamente descrita E194K. Esta mutación acabó reemplazándose en todos los virus G a lo largo de los pases de recuperación debido al cambio E196K. Probablemente, esta mutación haya condicionado bastante la evolución del linaje G, así como sus propiedades. Este cambio también ha sido relacionado con resistencia a AZT y a foscarnet entre otros antivirales, aunque es un polimorfismo frecuente (Tachedjian, French y Mills, 1998).

En su modelado se observó que de nuevo se sitúa en la anteriormente mencionada hélice α en el lateral de la palma (Figura 4.28). En este caso, produce una alteración opuesta de la carga electrostática de la zona, haciéndola mucho más polar (Figura 4.29). Probablemente el linaje G compensase esta carga con el efecto opuesto que producía la mutación anterior.

Al modelar la RT con la presencia de los dos cambios anteriormente descritos, E194K y G196E, y generar su superficie electrostática, se confirmaban sus efectos opuestos, observándose una recuperación de la neutralidad de la zona, similar a la del virus original (Figura 4.29).

Su eficacia biológica era menor que la del s61, aunque no llegaba a ser estadísticamente significativa por la variabilidad del experimento de competición (Figura 4.32). Esta mutación en la RT, al igual que las otras estudiadas en esta proteína, provocaba un aumento en producción de p24, que en este caso era el mayor de todos los observados. Su título viral era algo mayor que el del control y su citopatogenicidad algo menor.

El virus con el cambio G196E, al igual que los demás mutantes de RT, producía una disminución de la infectividad, ya que otra vez la razón título viral/p24 era mucho más bajo que el control. Todas las mutaciones en la RT mostraban como patrón común una baja eficacia unida a una baja infectividad.

4.6.1.2 Mutaciones aparecidas en los virus al realizar la semilla posterior a los pases placa a placa.

Se describen dos mutaciones que aparecieron en la región LTR de los virus al crecer las semillas derivadas de las placas después del Trinquete de Muller.

A586G (LTR)

Esta mutación apareció en los virus D en la región U5 de la LTR. Se fijó en todos los virus del linaje D al hacer la semilla para comenzar la recuperación de eficacia biológica (Figura 4.26). Este cambio revirtió en los virus D1p31 y D1.5p31, por lo que su efecto a priori era difícil de predecir.

Se modeló la estructura del ARN y se comparó con el s61, obteniéndose en este caso una estructura idéntica (Figura 4.27).

La mutación A586G tenía una eficacia similar a la del virus control (Figura 4.32). Sin embargo, este mutante producía mayor cantidad de defectivos, ya que tenía niveles muy altos de p24 junto con valores disminuidos de título viral. Era el mutante más citopatogénico de todos.

Probablemente, este virus se fijó al hacer la semilla del linaje D, debido a su rapidez en producir la muerte de los cultivos celulares y su elevada producción de partículas virales. Al realizar los pases de recuperación, sin embargo, esta mutación se ha revertido, hecho seguramente relacionado con su baja infectividad.

G587A (LTR)

La G en la posición 587 aparecía en la U5 de la LTR del virus original s61, mientras que los clones estudiados presentaban la A desde el principio. En éstos la G se adquirió al hacer la semilla a partir de los clones obtenidos tras el Trinquete de Muller (Figura 4.26). De este modo, se decidió introducir la adenina original presente en los clones (cambio G587A), para estudiar sus efectos sobre el virus y la causa de su sustitución al hacer las semillas de todos los virus estudiados.

En el modelado de su estructura, se pudo observar un ligero cambio en la estructura del ARN asociado a un aumento de energía libre estimada de sólo 2 kcal/mol, en concreto en la región de unión a cebador (Primer binding site en inglés (PBS)) (Beerens, Groot y Berkhout, 2001) (Figura 4.27).

Este cambio presentaba niveles significativamente menores de eficacia biológica que el control, niveles similares de p24 y niveles inferiores que el s61 de título viral. Por último, presentaba unos valores de viabilidad de los cultivos celulares igual que los del virus s61.

Esta mutación no supone ninguna ventaja para el virus y tiene peor eficacia, por lo que es normal que todos los virus adquiriesen la G en la posición 587 al hacer la semilla tras los pases en cuello de botella como paso inicial para recuperar eficacia.

4.6.1.3 Mutaciones aparecidas en los virus a lo largo de los pases de recuperación de eficacia biológica.

G593T (LTR)

Esta mutación se fijó en la U5 de la LTR en los primeros 10 pases de recuperación de los virus D (Figura 4.26). Esta mutación ya había sido descrita en un trabajo anterior llevado a cabo con estos mismos virus (Yuste y cols., 2005) y se mantuvo en los pases subsiguientes.

El cambio que producía en la estructura del ARN era igual que la G587A (LTR), modificando la zona del PBS (Figura 4.27).

Al analizar la eficacia de este mutante, se pudo observar que los valores eran similares al virus s61 (Figura 4.32). Lo mismo ocurría con el resto de características analizadas, detectándose niveles muy similares al control de producción de p24, título viral y citopatogenicidad, aunque con respecto a esta última presentaba una viabilidad algo menor que el virus s61.

Dados estos resultados y su recurrente aparición en virus del linaje D en distintos experimentos, lo más probable es que la T en la posición 593 tenga importancia en el proceso de recuperación de eficacia de estos virus.

V35I (p17)

Este cambio en la proteína p17 es uno de los más interesantes encontrados en todo este estudio, ya que era una mutación que aparecía en 6 virus de 3 linajes diferentes (G, I y K) (Figura 4.26). Esta mutación, como ya se indicó anteriormente, había sido descrita por su relación con la recuperación de eficacia de mutantes resistentes a inhibidores de la proteasa (Gatanaga y cols., 2002). El mencionado cambio se localiza en el extremo C terminal de una α -hélice, que corresponde con la señal de localización nuclear.

Se realizó un modelado por homología, de la p17 del s61 y de la proteína conteniendo esta mutación. Se comparó tanto la proteína simple como la acomplexada en dos subunidades con tres cadenas cada una, no pudiéndose encontrar cambios significativos en las estructuras obtenidas (Figura 4.30).

La I en la posición 35 de p17 suponía una eficacia algo mayor que la del virus s61, aunque esta diferencia no era estadísticamente significativa (Figura 4.32). Provocaba niveles más altos de proteína p24 y presentaba el título viral más alto de todos los mutantes, con valores de casi un logaritmo de diferencia con el control. Además, presentaba una citopatogenicidad mayor que el control, acabando con los cultivos ya en el día 5 post-infección.

Esta mutación provoca un aumento en todos los parámetros estudiados, aunque el de eficacia no era significativo. Este aumento era especialmente elevado en el título viral, indicando la gran infectividad de este mutante. Estos resultados demuestran la importancia de esta mutación en el proceso de recuperación y explican la adquisición de esta mutación por diferentes virus de linajes distintos.

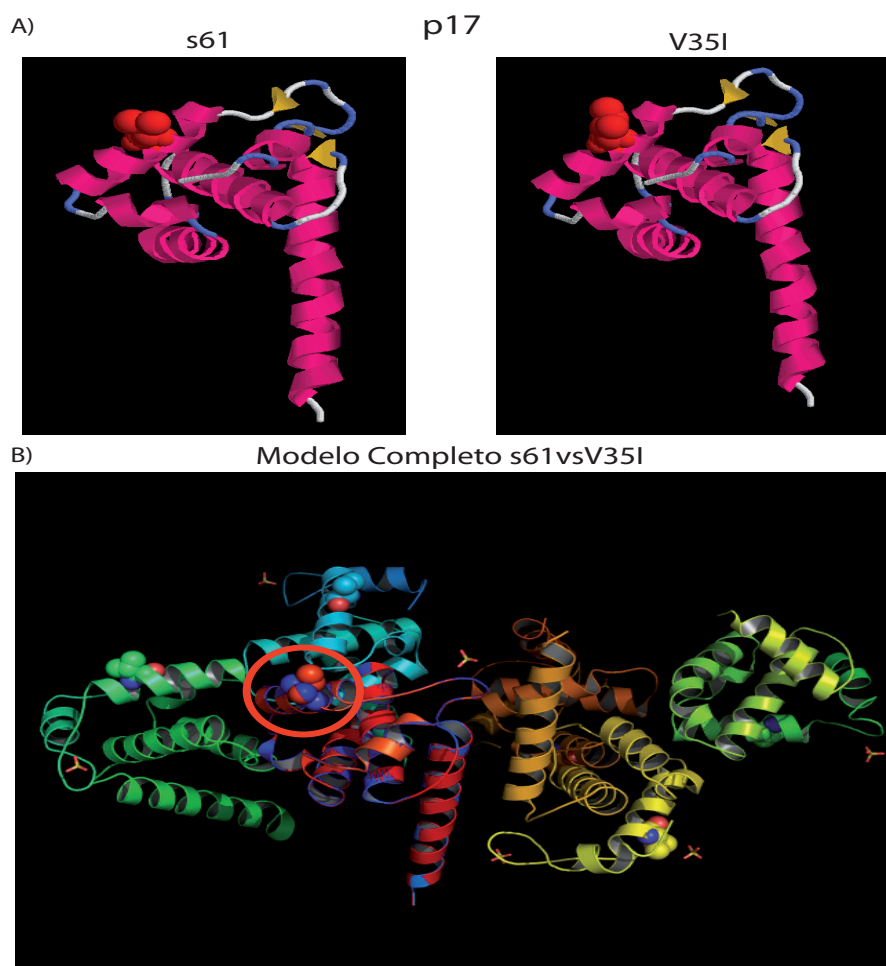


Figura 4.30 Modelado de la proteína p17 del virus s61 y del mutante V35I. A) Modelado de la proteína p17 de ambos virus en los que se muestran en rojo y con esferas el aminoácido presente en la posición 35. B) Modelo completo del complejo proteico que forma la proteína p17 con 6 proteínas acomplejadas en 2 unidades, con tres subunidades cada una. En una de las subunidades, se muestra una superposición de la proteína de s61 (azul) y la de V35I (rojo), con la posición 35 mostrada en esferas y señalada con el círculo rojo. En el resto de las subunidades se muestra la posición 35 salvaje con esferas para localizarla en el complejo proteico.

S314K (gp120)

El último cambio estudiado por mutagénesis fue el S314K en la región variable V3 de la proteína de la envuelta viral gp120. Este cambio aparecía en el pase 21 de recuperación de los virus G1 y G1.5, manteniéndose en el 31 (Figura 4.26). Implicaba dos cambios de nucleótido, G7170A y T7171A, que aparecían conjuntamente en dichos virus. En el virus G2p21 sólo se producía el segundo de los cambios llevando a un aminoácido diferente. Esta es una mutación importante que ha sido descrita en la literatura como la causante del cambio de tropismo celular R5 a X4. Al ser estos virus desde el inicio del tipo X4, el efecto de este cambio podría estar relacionado con su eficacia.

Al modelar esta proteína acomplejada con el receptor de linfocitos CD4, al que la gp120 se une naturalmente, se podía observar un cambio considerable producido por dicha mutación en la unión entre ambos (Figura 4.31). La presencia de la lisina producía un contacto mucho mayor entre el lazo V3 con el receptor CD4 que con la serina original.

Este mutante tenía un valor medio de eficacia del doble que el del virus control s61, sin embargo este aumento no era estadísticamente significativo debido a la gran variabilidad observada en el experimento de competición (Figura 4.32). Su producción de proteína p24 era similar a la del control, mientras que su título viral era mayor. Este virus era, además, más citopatogénico que el s61.

En resumen, este mutante tenía más eficacia que el control, mayor infectividad y citopatogenicidad, motivos por los que se habría fijado durante los pases de recuperación de los virus G1 y G1.5.

gp120

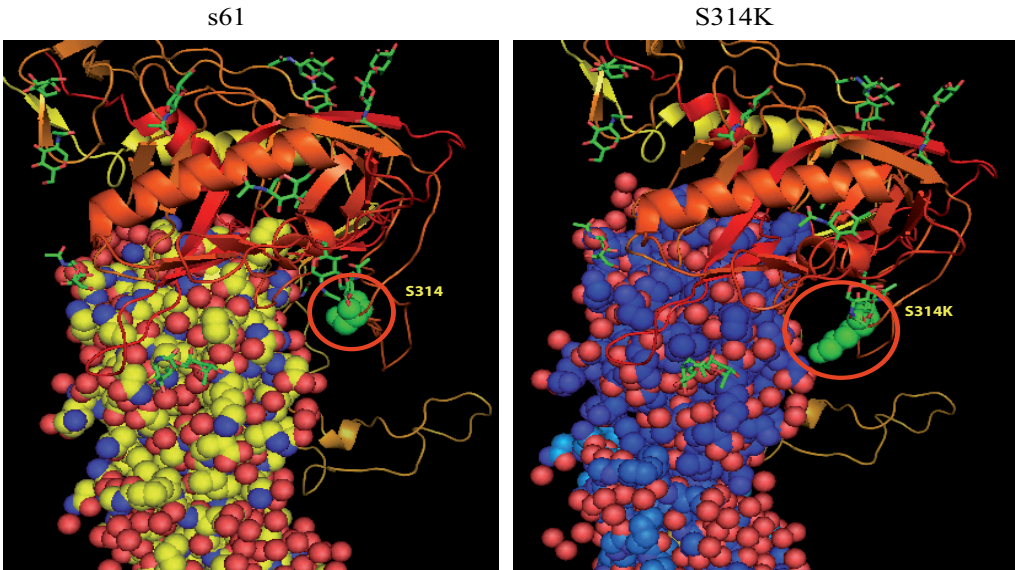


Figura 4.31 Modelado de la proteína gp120 del virus s61 y del mutante S314K. Ambas estructuras se muestran acomplejadas con el receptor celular CD4. Se muestra en cada modelo la posición 314 señalada con un círculo rojo, representándose el aminoácido presente en cada caso con esferas verdes.

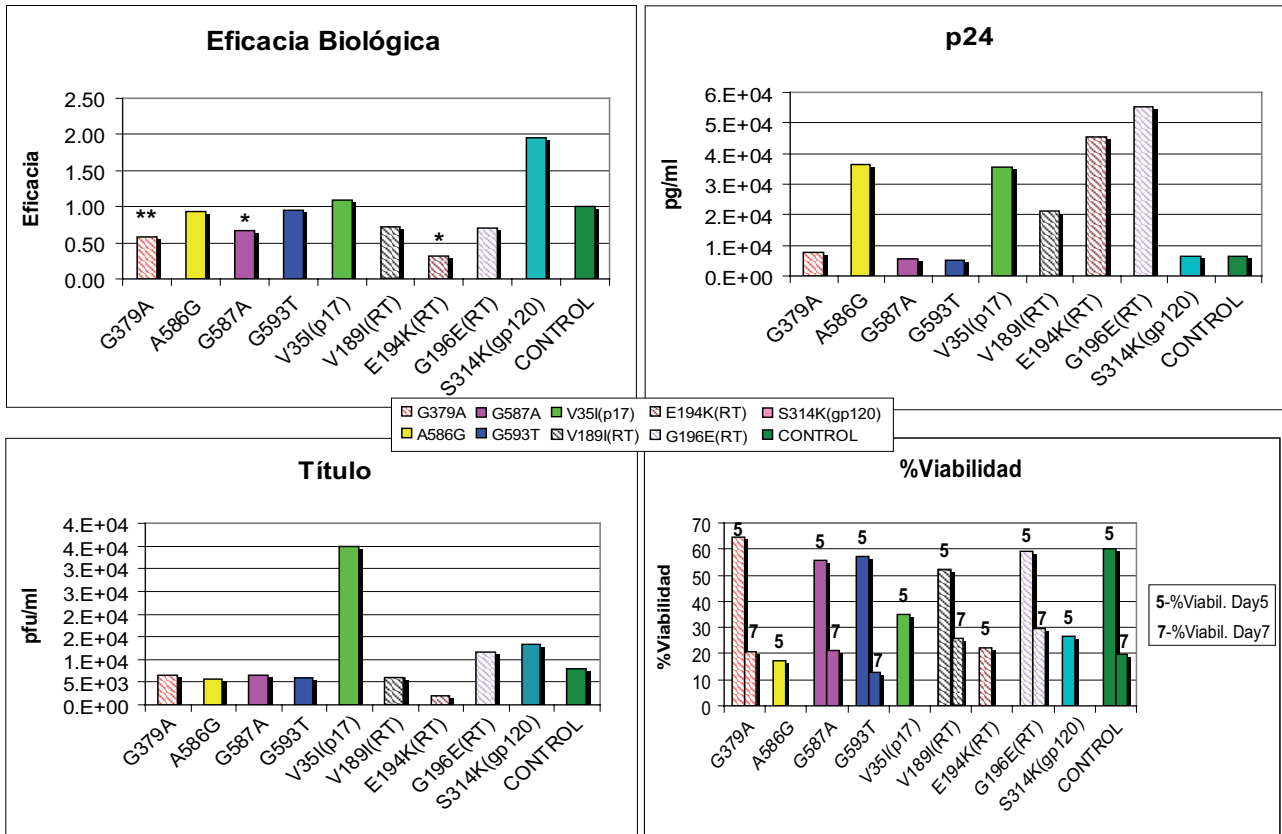


Figura 4.32 Caracterización fenotípica de los virus estudiados por mutagénesis dirigida. Se muestran los valores de eficacia biológica, producción de proteína p24, título viral y viabilidad de los cultivos en el ensayo de citopatogenicidad. Los datos de cada virus se muestran del mismo color según la leyenda mostrada. Los virus con mutaciones aparecidas durante los pases placa a placa aparecen además rayados. El virus s61 clonado constituye el control de todos los experimentos. En los datos de eficacia, se muestran los virus con diferencias estadísticamente significativas con respecto al control con asteriscos (* = p-valor<0,05 y ** = p-valor<0,01). En el ensayo de citopatogenicidad, se muestran los porcentajes de viabilidad de los cultivos tras 5 y 7 días post-infección (excepto los virus que acabaron con el cultivo antes en los que solo se muestra la viabilidad a día 5).

4.6.2 Resumen de los resultados obtenidos en el estudio de mutantes.

Después de haber analizado las propiedades de los mutantes elegidos, los resultados generales obtenidos se podían resumir como:

- Las mutaciones aparecidas durante los pases seriados en cuello de botella suponían un importante coste en la eficacia del virus, como predice el Trinquete de Muller.
- Las mutaciones aparecidas en los cuellos de botella, así como algunas presentes en los clones iniciales, condicionaron la posterior evolución de los virus. Así, la mutación G379A (LTR) ha llevado a los virus D a acumular mutaciones compensatorias en la LTR, la presencia de G196E (RT) en los virus G ha hecho que estos virus acumulen una mutación a priori perjudicial para compensar la carga de la zona o G587A (LTR), presente en todos los virus, se revertía en la semilla como forma de recuperación inicial.
- Las mutaciones en la RT aparecidas durante los pases placa a placa, produjeron una caída fuerte en la infectividad de los virus y su efecto en solitario era perjudicial para los virus.
- La acumulación de mutaciones en la LTR al realizar la semilla de los clones tras los pases placa a placa, indica que mutaciones en esta región pueden suponer aumentos rápidos de eficacia.
- Las mutaciones fijadas en la recuperación y mantenidas en el proceso suponen mejoras en la infectividad y en la citopatogenicidad, pero no una mejora significativa en la eficacia.
- La mutación V35I (p17) aumentaba todos los parámetros estudiados con valores algo mayores de eficacia que el s61, por lo que se comprende que sea convergente en tres linajes diferentes.
- Algunas mutaciones se han fijado durante la recuperación por su mayor velocidad en la infección, pero luego han revertido por su déficit en otras características. Ejemplos de esto son las mutaciones A586G (LTR) o E194K (RT) (esta última siempre teniendo en cuenta que compensaba la G196E (RT) en los virus G). Estos virus mostraban una alta citopatogenicidad, pero presentaban valores bajos en título y eficacia biológica.
- El virus es capaz de superar los problemas surgidos de la acumulación de mutaciones deletéreas de muy diversas formas, como demuestra la recuperación de los virus estudiados. Sin embargo, el efecto individual de las mutaciones fijadas en los pases de recuperación, no explica las fuertes subidas de eficacia observadas en el proceso.

4.7 Generación de un paisaje de eficacias a partir de datos reales y su aplicación para el estudio del proceso de recuperación de eficacia del VIH-1 y la estimación de eficacias desconocidas.

4.7.1 Generación de un paisaje de eficacias reales utilizando métodos filogenéticos.

Después de los diversos experimentos llevados a cabo, se contaba con una serie de virus estudiados durante la pérdida y recuperación de eficacia “in vitro” del VIH-1. Estos virus habían sido obtenidos después de los ya descritos pases placa a placa y en gran tamaño poblacional (Materiales y Métodos, Figura 3.1). De dichos virus se conocían muchas de sus características virológicas (título, p24, actividad RT...) y, sobre todo, sus secuencias y eficacias biológicas en cada momento. Por lo tanto, se decidió utilizar los mencionados datos para representar, por primera vez, un paisaje de eficacias de la recuperación de la eficacia “in vitro” del VIH-1. Esta representación no sería teórica, ni utilizaría eficacias estimadas, sino que se basaría exclusivamente en datos experimentales.

Para ello, se utilizaron todas las secuencias y datos de eficacia de los que se disponía, desde los clones iniciales, pasando por los pases en cuellos de botella, hasta los pases en gran tamaño poblacional. En total, se disponía de 55 secuencias de 9451 nucleótidos cada una. Una vez seleccionados todos los virus, se segregaron utilizando diversos métodos filogenéticos, entre los que estaban Neighbour Joining, Máxima Verosimilitud o Redes de Medianas. Todos ellos mantenían los diferentes linajes separados y agrupados entre sí, al igual que pasaba con los resultados filogenéticos previamente expuestos, y desplegaban topologías muy similares. Los resultados que se muestran son los obtenidos con Máxima Verosimilitud (Figura 4.33A). Con esta separación, podíamos disponer de valores de distancias genéticas, que podían ser utilizadas como coordenadas bidimensionales para representar estos virus en un plano XY. Con la adición de los valores de eficacia asociados a cada secuencia global, ya se tenía la tercera dimensión necesaria para dar altura a los valores y representarlos en un eje XYZ. La posterior utilización del programa SurGe de generación de imágenes matriciales, en el que se generaba una matriz de puntos utilizando los datos XYZ, permitía representar el paisaje generado a partir de los datos reales de los que se disponía, como se muestra en la Figura 4.33B.

El paisaje obtenido mostraba una topología rugosa y, en dicho paisaje, se podía describir cómo había sido el proceso de pérdida y recuperación de eficacia en clones del VIH-1. En él se observaban zonas más continuas y otras mucho más abruptas y, por lo tanto más proclives a subidas y bajadas drásticas de la eficacia. Se obtuvo un pico máximo correspondiente al clon inicial I1, que era el de mayor eficacia al principio. Cerca de este pico, en una región más plana, se situaba el virus original s61. Los virus se situaban en diferentes regiones del paisaje, habiendo virus en zonas más llanas, como los virus del linaje E y H, en los que el incremento de eficacia era constante y sin aparentes riesgos de pérdida de eficacia, mientras que otros virus, como los pertenecientes al linaje D o K, se situaban en zonas muy abruptas, cerca de estrechos picos y valles. Estos últimos virus corrían riesgos mayores de extinción o disminución de eficacia en su proceso de recuperación.

Esta metodología ilustraba perfectamente el proceso evolutivo ya ocurrido, pero carecía de poder predictivo, debido a que los métodos filogenéticos no producen topologías fijas. Las distancias en los árboles filogenéticos son unidireccionales y sólo tienen valores reales a lo largo de las ramas. Por lo tanto, los métodos filogenéticos no otorgan valor a las distancias laterales entre las secuencias analizadas, es decir, un árbol se puede voltear sin modificar su significado. Además, a medida que se añaden más secuencias a un árbol filogenético, éste añade ramas nuevas para incluirlas, cambiando, a su vez su topología. Estos problemas no son importantes para describir un proceso evolutivo, pero sí que lo son para estimar valores desconocidos, donde se requiere una topología inmutable, por lo que esta metodología no permitía la posibilidad de estimación de valores desconocidos.

A)



B)

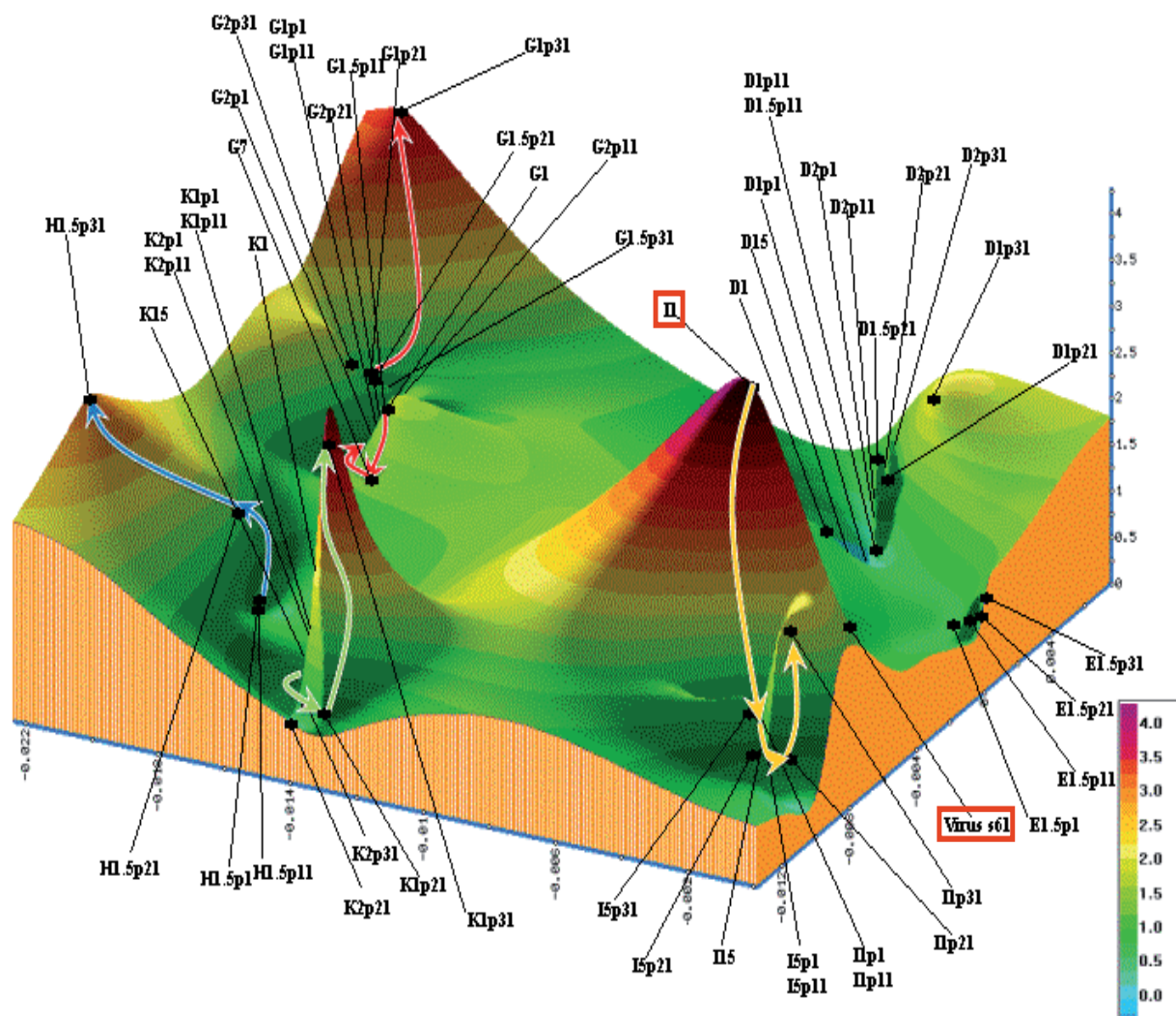


Figura 4.33 Paisaje de eficacias de los virus recuperados por métodos filogenéticos. Se utilizaron las secuencias completas de todos los virus sometidos a pases placa a placa y a pases seriados en gran tamaño poblacional. A) Las secuencias de los genomas completos de todos los virus estudiados (55 en total) fueron separadas filogenéticamente utilizando el método de Máxima Verosimilitud. El árbol obtenido representa la distancia genética entre los diferentes virus, expresada como número de sustituciones por sitio. B) Tras obtener el árbol, se utilizaron las distancias genéticas como coordenadas X e Y. A estos puntos se les aplicaron los valores de eficacia calculados experimentalmente como valores Z. Tras aplicar el programa de rasterizado SurGe, se obtuvo una representación tridimensional del paisaje de eficacias correspondiente a la evolución “in vitro” de los virus de VIH-1 analizados. La posición de cada virus en el paisaje obtenido se indica en la figura. La escala de colores representa los valores de eficacia biológica, siendo el azul oscuro el valor más bajo y el rojo oscuro el más alto. Las flechas representan la trayectoria seguida por varios de los virus estudiados como ejemplo de los caminos evolutivos observados. La localización en el paisaje del virus original s61 y del clon inicial I1, con el valor más alto de eficacia biológica, se indican con un recuadro rojo.

4.7.2 Generación de un paisaje de eficacias real predictivo utilizando redes neuronales artificiales.

Para solventar los citados problemas para la predicción de valores de eficacia desconocidos, se decidió hacer uso de métodos de redes neuronales artificiales en colaboración con el Prof. Morán de la Universidad Complutense de Madrid y la Prof. Delgado de la Universidad Politécnica de Madrid. En particular, se utilizaron las SOM desarrolladas por Kohonen, explicadas en Materiales y Métodos. Este método mapea datos multi-dimensionales sobre una superficie bidimensional, basándose en distancias Euclídeas. Al tratarse de secuencias, fue necesario hacer previamente una codificación de dichas secuencias nucleotídicas utilizando un procedimiento de tetraedro irregular (ver Materiales y Métodos, apartado 2.9.2). Tras varias rondas de aprendizaje, las distintas secuencias quedaban asociadas a un nodo de la red creada y, a cada nodo una vez etiquetado con la secuencia correspondiente, ya se le podía asignar su valor experimental de eficacia, para dotar a la representación de una tercera dimensión. El resto del paisaje se generó etiquetando a cada vector con la media de los valores de los L vectores más próximos ($L=1$). Al haber segregado los virus sobre una superficie con un tamaño fijo y preestablecido, todas las distancias obtenidas entre secuencias son reales y fijas. La adición de nuevas secuencias a los mapas obtenidos no variaría ni la disposición de los virus previamente añadidos, ni la topología. Con esta nueva representación es posible tener capacidad de predicción y estimación.

En primer lugar, esta metodología se aplicó a las mismas secuencias globales de los virus completos que se utilizaron para la representación anterior. El objetivo inicial era comparar ambas representaciones y analizar los resultados obtenidos. Además, con este nuevo método se podía calcular el error medio de eficacia estimada asociado a esta representación. Con las 55 secuencias de los genomas completos se obtuvo el paisaje de eficacias del proceso de recuperación (Figura 4.34A). La topología del paisaje era muy similar a la previamente obtenida con los métodos filogenéticos, lo que apoyaba el uso de la nueva metodología. Como se puede observar en la Matriz U (Figura 4.33B), en la que se muestra la distancia entre cada nodo, los virus se agrupaban correctamente según su genealogía. El virus original volvía a aparecer cerca del pico máximo de I_1 y cada virus aparecía en regiones continuas o abruptas igual que antes. El examen de las trayectorias seguidas por cada virus (Figura 4.34C), mostraba la convergencia de los distintos virus del mismo linaje. Esta representación permitía un análisis muy claro de la trayectoria evolutiva de los diferentes virus estudiados. Por último y debido a las características del sistema, se pudo calcular el error medio del método. Para ello, se volvió a entrenar la red quitándole uno de los virus del conjunto y estimando luego su eficacia para determinar la diferencia entre el valor estimado y el real. Repitiendo este proceso de manera reiterada para todos los virus utilizados. El resultado obtenido fue un error medio del 11.6% con respecto al valor real de eficacia, resultado muy razonable teniendo en cuenta la gran variabilidad de los datos de eficacia que se obtienen en los experimentos de competición.

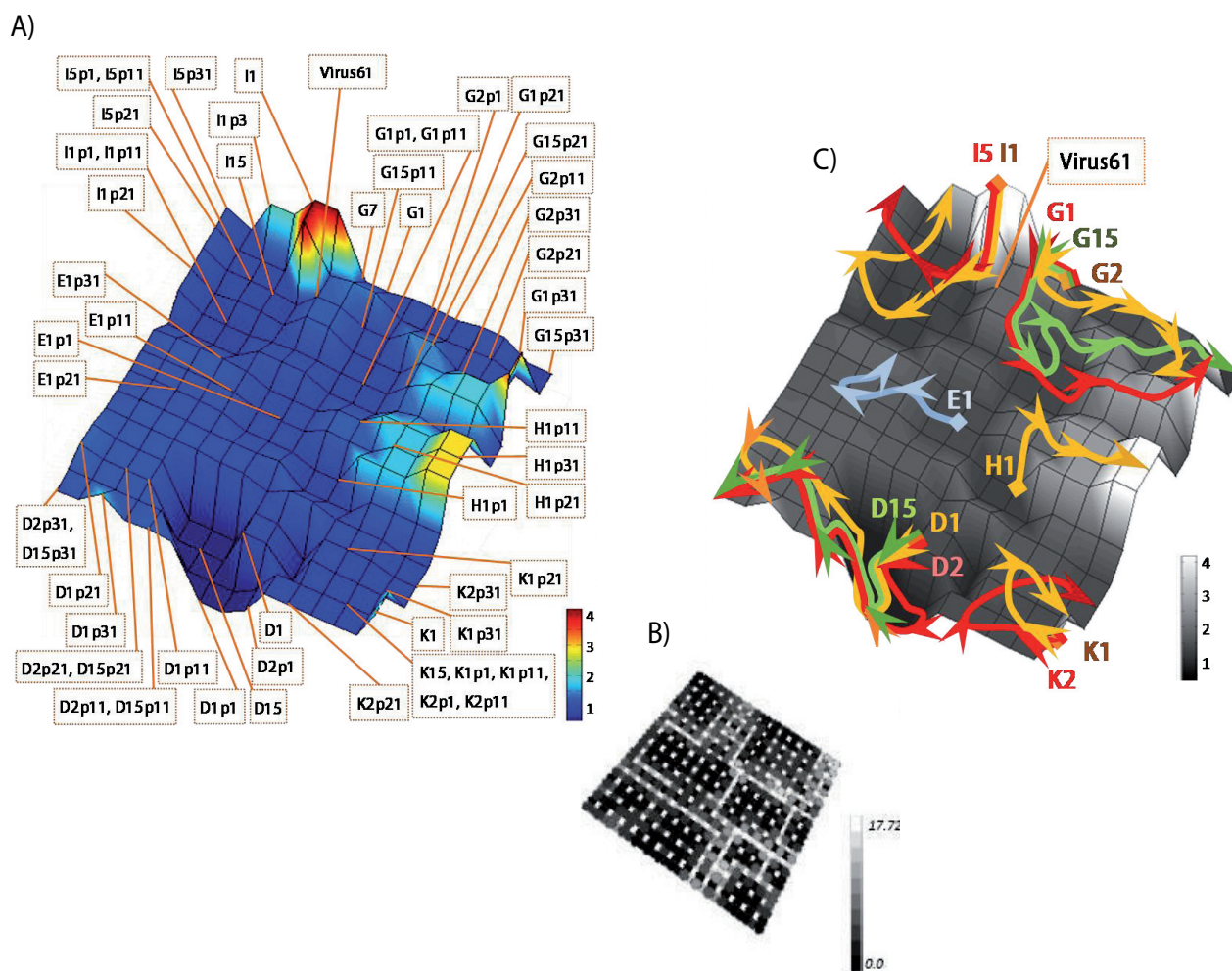


Figura 4.34 Paisaje de eficacias de los virus estudiados utilizando redes neuronales artificiales. Una red SOM bidimensional de 15 x 15 fue entrenada con las secuencias del genoma completo de los 55 virus de VIH-1 estudiados. El proceso de aprendizaje se llevó a cabo como se explica en Materiales y Métodos. A) Paisaje de eficacias obtenido mediante SOM etiquetado con los nombres de cada virus. La escala de color representa los diferentes valores de eficacia biológica. B) Matriz U de las distancias euclídeas entre cada uno de los nodos del mapa SOM obtenido. Esta matriz representa las distancias reales entre los nodos de la red obtenida con las secuencias utilizadas. La escala muestra la distancia entre nodos, siendo el negro la distancia más corta y blanco la más larga. Se pueden observar los grupos de secuencias obtenidos separados entre ellos por regiones blancas (de distancias mayores). C) Sobre el mismo mapa mostrado en (A) las flechas indican los caminos evolutivos seguidos por los diferentes virus durante los pases llevados a cabo. En este caso la escala de negro a blanco es la que indica los valores de eficacia biológica.

Una vez comprobada la validez del método, se procedió a realizar la estimación de eficacias desconocidas. En este caso, se pretendía estimar las eficacias desconocidas de las distintas variantes virales presentes en las cuasiespecies de los virus estudiados. Se utilizaron las cuasiespecies de las que ya se disponía en la región V1-V2 de la envuelta. En primer lugar, se generó de nuevo el paisaje de eficacias de la recuperación de los virus, pero esta vez utilizando exclusivamente las secuencias consenso de esta región del mismo grupo de virus. Este grupo consistía en 55 secuencias de 527 nucleótidos a las que se podía asignar los valores de eficacia experimentales previamente utilizados. El paisaje obtenido tenía características topológicas similares al paisaje de genomas completos, pero con mucha menor resolución, a causa de la disminución de la información debido a la menor longitud de las secuencias (Figura 4.35A). Una vez obtenido el paisaje, se procedió a medir su error medio de igual manera que con el de los genomas completos. El valor calculado fue del 10,9%, muy parecido al anteriormente obtenido.

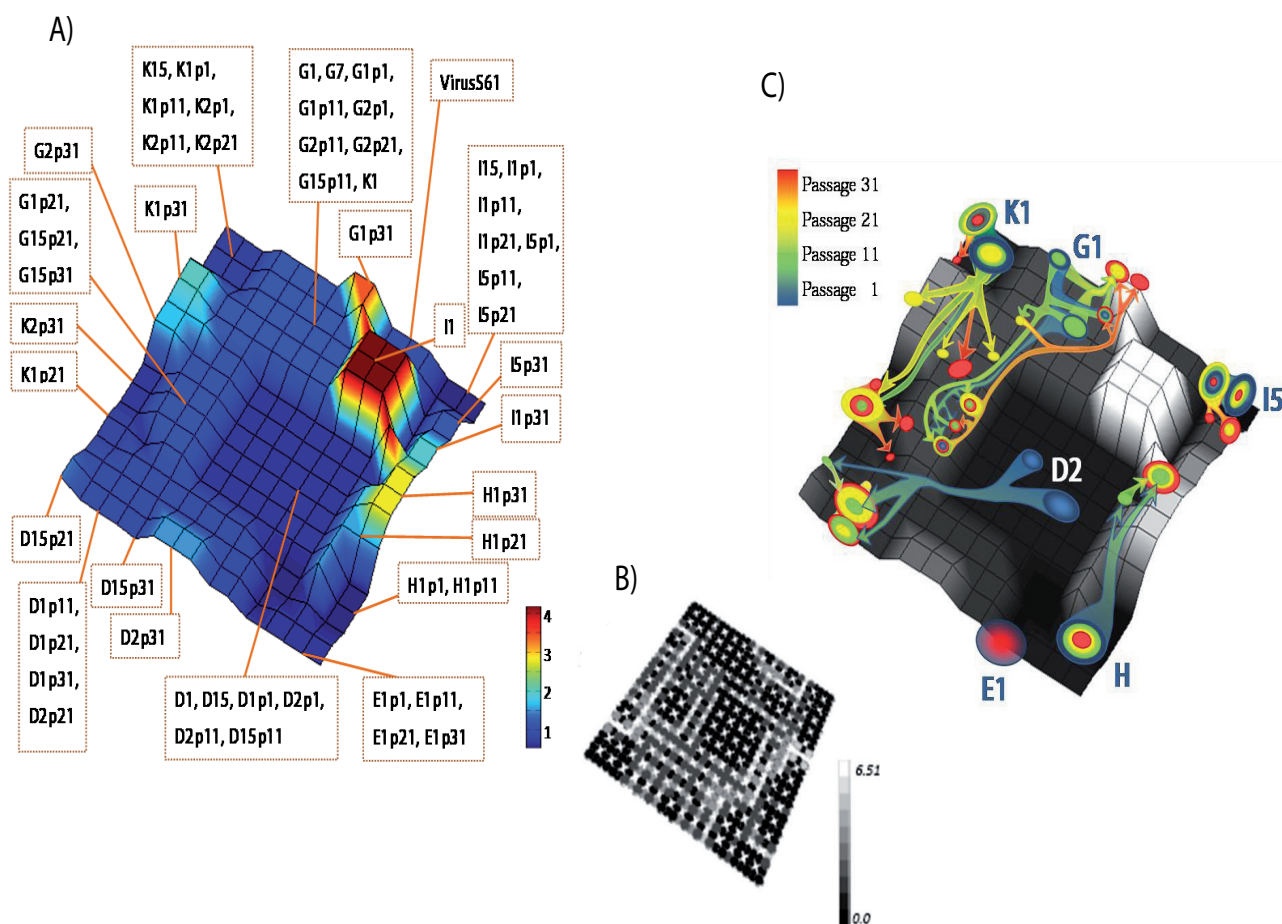


Figura 4.35 Paisaje de eficacias de las secuencias nucleotídicas de la región V1-V2 del gen *env* utilizando redes neuronales artificiales. Se entrenó una red SOM de 15 x 15 con las 55 secuencias consenso de V1-V2, que posteriormente se utilizó para estimar la eficacia de los clones obtenidos de cada virus en dicha región a lo largo de los pases de recuperación de eficacia biológica (911 secuencias). A) Paisaje de eficacias obtenido de la región V1-V2 del gen *env* con la localización de cada virus indicada por sus nombres. La escala de colores representa la eficacia biológica. B) Matriz U del paisaje de V1-V2 indicando la distancia entre cada nodo. C) Proyección de las 911 secuencias de las cuasiespecies de cada virus con eficacia desconocida. Los valores de eficacia de cada clon se estimaron utilizando el mapa obtenido con las secuencias consenso. Las flechas muestran como ejemplo la evolución de las cuasiespecies de varios de los virus (D2, G1, I5 y K1) a lo largo de los pases de recuperación. La transición de colores en cada flecha indica la evolución desde un pase al siguiente. En la escala se muestra el color correspondiente a cada pase.

A partir de este paisaje, se añadieron 911 secuencias de 527 nucleótidos disponibles, correspondientes a las cuasiespecies en la región V1-V2 de *env* de cada virus en cada pase. Como ya se tenía el paisaje de la envuelta etiquetado, únicamente era necesario proyectar la secuencia de cada clon en el mapa y asignarle la eficacia del nodo que más se aproximase. De esta forma se consiguió tener para cada clon un valor de eficacia asociado. El análisis de estas eficacias estimadas permitió observar la evolución de las cuasiespecies a lo largo de los pases de recuperación (Figura 4.35C). Se pueden observar caminos constantes de subida como el del virus H, aunque mantenía siempre una representación minoritaria de variantes de la cuasiespecie inicial. Otros virus realizaban una compleja búsqueda hasta alcanzar un pico de eficacia (virus G1 y K1) o recorrían grandes distancias (D2). Por último, otros virus no cambiaban en esta región y se mantenían en el mismo punto del paisaje, acumulando simplemente variabilidad (virus E). Con esta metodología, además, se podía estimar razonablemente cómo iba a evolucionar la cuasiespecie en pases posteriores, ya que los clones generados en un pase con mayor eficacia estimada, se imponían en los siguientes. Así por ejemplo, en

el linaje H (Figura 4.35C) se puede observar como en el pase 11 había clones en 2 puntos nuevos del paisaje, ambos con mayor eficacia estimada que el inicial que era de 0,51, uno con valor estimado de 1,57 y el otro con 2,82. Al continuar con los pases, se comprueba que la población casi al completo se traslada al punto que tenía mayor eficacia estimada, el de 2,82, abandonando el otro punto explorado de menor eficacia. Esto concuerda con lo esperado, repitiéndose en las demás cuasiespecies, por lo que se puede predecir de manera razonable la evolución que puede seguir la población viral en pases posteriores. Si existen clones con eficacias mayores, éstos se impondrán en la población, mientras que si los valores estimados de los clones no superan al consenso actual, esta población no se moverá.

Con este método, se ha conseguido, no sólo representar por primera vez un paisaje de eficacias real del proceso de recuperación “in vitro”, sino también generar una herramienta de estimación de eficacias desconocidas asociadas a secuencias relacionadas.

5. Discusión

El trabajo realizado en la presente tesis constituye un extenso estudio del proceso de recuperación de eficacia biológica “in vitro” del VIH-1. En este estudio se ha analizado dicho proceso utilizando diversas metodologías y abordajes, todos ellos complementarios, para comprender globalmente los mecanismos que sigue el VIH-1 con el fin de recuperar eficacia después de una pérdida drástica de la misma, debida a severos cuellos de botella poblacionales.

El estudio de este proceso resulta muy útil para entender la evolución natural del VIH-1, ya que generalmente toda infección comienza con un fuerte cuello de botella de la población viral (Salazar-Gonzalez y cols., 2009). La disminución de variabilidad que provocan estos cuellos de botella resulta perjudicial para el virus, debido a que dicha variabilidad, como se ha podido comprobar en esta tesis, es uno de los factores fundamentales para la eficacia viral. Después de esta drástica disminución del tamaño poblacional con la que se inicia cada infección, se desarrollan en el virus mecanismos para recuperar la eficacia perdida y, de esta manera, continuar la infección del huésped con éxito. Por estos motivos, en este trabajo se ha intentado simular dicho proceso de pérdida y recuperación de eficacia en la infección natural, utilizando para ello técnicas de cultivo “in vitro” del virus. Utilizando este tipo de experimentos, hemos podido conocer gran número de características de los virus en recuperación y examinar la importancia que tiene cada una de ellas en la recuperación de la eficacia biológica. Se ha intentado, además, elaborar una hipótesis general de cómo transcurre el proceso de recuperación y sus características principales.

Los resultados obtenidos en todos los experimentos realizados, sugieren que la producción de variabilidad en las cuasiespecies virales supone el mecanismo fundamental para la recuperación de eficacia biológica “in vitro”, como demuestra la fuerte correlación encontrada entre la heterogeneidad de las cuasiespecies y los valores de eficacia (Figura 4.9). Este proceso de recuperación, parece completamente dirigido por la selección positiva, debido a la gran preponderancia de mutaciones no sinónimas, no sólo a nivel de genoma completo, sino a nivel del espectro de mutantes que compone la población viral. Esta generación de variabilidad permite al virus explorar mayor espacio en el paisaje de eficacias y encontrar multitud de soluciones evolutivas diferentes para mejorar su eficacia biológica.

5.1 Análisis de la recuperación de eficacia biológica.

En el estudio realizado, se ha observado que, después de 30 pases en gran tamaño poblacional, la eficacia biológica sigue aumentando globalmente no habiéndose llegado aún a una estabilización de los valores. A pesar de la variabilidad de comportamiento observada en la recuperación de los distintos virus, la mayoría ha aumentado su eficacia a lo largo de los pases en gran tamaño poblacional.

En este análisis de la recuperación de eficacia, llama poderosamente la atención la existencia de dos virus (D1.5 y G1.5) que han perdido eficacia biológica en los últimos 10 pases. Los mencionados virus D1.5 y G1.5, han sufrido caídas de eficacia tras haber sido dos de los virus con mayor eficacia en pases anteriores; así, en el pase 21 eran el segundo y tercer virus con mayores valores. Estos virus presentaban características de alta adaptación en el pase 21, como demuestran los niveles más altos

de título viral (ambos virus con un valor de $1,78 \times 10^6$ TCID₅₀/ml frente a un valor medio del pase 21 de $6,86 \times 10^5$) y actividad RT (con valores de $1,33 \times 10^5$ PSL para D1.5p21 y $1,73 \times 10^5$ para G1.5p21 frente a una media del pase 21 de $6,37 \times 10^4$ PSL), pero con valores bajos de heterogeneidad en sus cuasiespecies (indicados en Figura 4.9). Este resultado contrasta con lo comúnmente aceptado en este tipo de estudios con otros virus como el VSV o el VFA, en los que se observa un aumento continuo de la eficacia biológica al realizar pases seriados, hasta que se alcanza una estabilización de los valores, a partir de la cuál la eficacia comienza a fluctuar en torno a los valores alcanzados (Escarmis, Davila y Domingo, 1999; Novella y cols., 1995).

Durante todo el estudio se ha observado que el cultivo del VIH-1 en gran tamaño poblacional no supone un entorno selectivo constante, sino que supone un entorno dinámico con una evolución continua. Esto lo demuestra la gran diversidad de soluciones evolutivas observadas durante los pases (Figura 4.4), e incluso la distinta localización de las mutaciones aparecidas entre los diferentes pases (Figura 4.6). Por este motivo, los virus que han comprometido la heterogeneidad de sus cuasiespecies en favor de una alta adaptación, han perdido eficacia al continuar los pases por su falta de capacidad adaptativa. Esto es lo que se ha observado en los virus D1.5 y G1.5, en los que, probablemente por la selección de variantes muy eficaces en un momento determinado, se ha producido un cuello de botella poblacional. En ambos casos, los virus que perdían su eficacia eran las réplicas realizadas en 5×10^6 células, mientras que las realizadas en $2,5 \times 10^6$ continuaban aumentando sus valores. Se puede suponer que los virus D1.5 y G1.5 al haber sido cultivados en un tamaño de población mayor que sus réplicas D1 y G1, han tenido más posibilidades de fijar con más rapidez un mayor número de mutaciones con alta eficacia asociada, que, al imponerse, han producido los cuellos de botella en la población viral. A pesar de su pérdida de eficacia y la gran delección que ha fijado uno de ellos, el virus G1.5 (Figura 4.4), estos virus son perfectamente viables, presentando una citopatogenicidad algo mayor que la media en el pase 31, con unos valores de viabilidad de los cultivos al tercer día post-infección de 20% y 29% para D1.5p31 y G1.5p31, respectivamente, frente a una viabilidad media en el pase 31 de 36%. Esto quiere decir que estos virus no han perdido infectividad, sino que la pérdida de eficacia observada unida a la pérdida de heterogeneidad, sugieren una merma en su capacidad adaptativa. Este resultado no hace sino confirmar la importancia, observada a lo largo de todo el proceso, de la heterogeneidad de las cuasiespecies virales.

El resto de los virus muestran una gran variabilidad de comportamiento, aunque siguen un modelo esperable, con aumentos de eficacia cuando los virus son capaces de salir del “agujero” de eficacia del que partían tras el Trinquete de Muller. Incluso, uno de los virus estudiados, el K2, hasta el momento ha sido incapaz de recuperarse del proceso del Trinquete de Muller, probablemente porque no ha podido contrarrestar los efectos deletéreos de las mutaciones fijadas durante los cuellos de botella, especialmente la mutación R15K en la proteína p17, como indica el análisis de redes mostrado en la Figura 4.19, y la E194K en la RT, cuyo coste de eficacia se comprobó en el análisis mediante mutagénesis dirigida (apartado 4.6.1.1). Excepto en este caso, el resto de los virus ha sido capaz de recuperar su eficacia, pero presentando cada uno de ellos diferentes patrones de recuperación. Los virus, en general, han mostrado una adaptación secuencial, como demuestran los valores de eficacia de la Figura 4.2, aumentando su eficacia una vez encontradas las soluciones evolutivas adecuadas en cada caso.

5.2 Análisis genético del proceso de recuperación de eficacia biológica.

Los genomas de los virus obtenidos durante los pases de recuperación han sido analizados exhaustivamente, tanto en sus secuencias globales, como en las cuasiespecies virales que conformaban las poblaciones.

El primer estudio se llevo a cabo analizando las secuencias completas de los virus a lo largo de los pases en gran tamaño poblacional. En dichos análisis se pudieron detectar las mutaciones impuestas en cada pase y su dinámica a lo largo de la recuperación. En los primeros 10 pases (hasta el pase 11), ya se había observado previamente que el número de mutaciones fijadas era muy escaso, incluso con la presencia de virus que no acumulaban ninguna mutación, lo que indicaba que los virus pueden ganar eficacia sin cambiar la secuencia consenso (Borderia y cols., 2010). A pesar de lo escaso de estas primeras mutaciones, los resultados obtenidos con las secuencias de pases posteriores indicaban la enorme importancia de estos cambios en la recuperación de eficacia, ya que más del 80% se han mantenido fijadas a lo largo de todos los pases (véase Figura 4.5). Estas primeras mutaciones parecen constituir la base de la recuperación de los virus que las adquirían y han condicionado su evolución posterior. A partir de esta base genética generada en los primeros pases, los virus han continuado explorando el paisaje de secuencias fijando más mutaciones en los pases subsiguientes (Tabla 4.1). Estas mutaciones fijadas posteriormente, aunque algunas con importantes efectos fenotípicos, son más un reflejo del aumento de la complejidad de las poblaciones virales, como demuestran el mantenimiento en la población del pase 31 de tan sólo el 26% de las mutaciones fijadas en el pase 21 y el gran aumento de posiciones heterogéneas. Este aumento de la complejidad de la población provoca que se observen mutaciones en los genomas consenso del pase 21 que posteriormente se han remplazado de nuevo por el nucleótido original presente en los virus iniciales. Esto ocurría debido a que dichos cambios no se habían impuesto completamente en la población, sino que diversas variantes convivían en el momento de la aparición de la mutación observada en la secuencia consenso. Por lo tanto, aunque el número de mutaciones aparecidas en los genomas completos a partir del pase 21 aumentó de forma considerable, seguía sin apreciarse una relación directa entre los cambios aparecidos y el aumento de eficacia observado. Así, el incremento en las mutaciones fijadas no ha seguido el mismo comportamiento que el incremento de eficacia biológica (compárese Figura 4.1 con Figura 4.5A).

A pesar de la gran variabilidad de mutaciones aparecidas en los genomas completos de los diferentes virus, se pudo observar la presencia de cambios comunes en varios virus. Se observaron 9 eventos de convergencia, 8 de ellos en virus del mismo linaje y uno en virus de linajes diferentes (la mutación V35I de la p17 que aparecía en los linajes G, I y K). Estas convergencias indicaban la existencia caminos evolutivos comunes mediante mutaciones recurrentes para la recuperación de la eficacia. Además, se observó una acumulación de deleciones en la región variable V1-V2 de la envuelta del virus en la mayoría de los virus estudiados (Figura 4.7). Este hecho ya había sido descrito anteriormente en estudios con pacientes, en los que se observó que virus con regiones variables más cortas eran más eficaces en la transmisión (Derdeyn y cols., 2004). La acumulación de deleciones en esta región en pases “in vitro” demuestra que es un recurso habitual para la mejora de eficacia. Esto se debería a que la disminución del tamaño del lazo estructural que forma esta región facilita la unión

entre la gp120 y el receptor CD4.

Debido a que aun no se encontraba una relación directa entre las mutaciones en los genomas globales y la recuperación de eficacia biológica observada, se decidió ir más allá y estudiar el espectro de mutantes de los virus en los diferentes pases. Para ello, se tuvo en cuenta el modelo de cuasiespecies y estudios previos que demostraban efectos fenotípicos de la composición de las cuasiespecies (Escarmis y cols., 2008; Vignuzzi y cols., 2006). Así, se calculó la heterogeneidad de las cuasiespecies en cuatro regiones diferentes del genoma de cada virus en los diferentes pases analizados. Las distintas regiones analizadas se eligieron para tener una representación de distintas regiones genómicas con diferentes funciones. La región de la LTR se eligió por ser una zona no codificante y estar relacionada con la replicación, la zona de p17 de *gag* por codificar una proteína estructural, *vpu* por ser un gen accesorio y las dos regiones variables V1-V2 y V3-V4 de *env*, por ser regiones muy relacionadas con la variabilidad del virus. Utilizando este análisis encontramos una correlación muy significativa entre la heterogeneidad de las cuasiespecies virales y la eficacia biológica (Figura 4.9).

En contra de este resultado se podría argüir que la heterogeneidad sólo estaría indirectamente relacionada con el incremento de eficacia por la mera acumulación de mutaciones al azar durante los pases seriados. Los virus irían acumulando cambios a medida que se realizasen los pases, por lo que se vería un aumento de la heterogeneidad, y a la vez estos virus incrementarían su eficacia por otros motivos. Sin embargo, varios de los resultados obtenidos eliminan esta posibilidad, afianzando la hipótesis inicial de la relación entre heterogeneidad y eficacia. En primer lugar, los ya mencionados casos de los virus D1.5 y G1.5. Estos virus que cayeron de eficacia presentaban valores disminuidos de heterogeneidad, lo que no hubiese ocurrido con una acumulación constante de mutaciones. Además, el caso de estos virus mejoraba la correlación obtenida, ya que su baja heterogeneidad y eficacia caían sobre la curva obtenida, mejorando el *p-valor* de la estimación.

Por otro lado, la gran preponderancia de mutaciones no sinónimas dentro de las cuasiespecies también está en contra de la acumulación constante de mutaciones al azar, ya que si esto estuviese ocurriendo no se observaría la acumulación de dN sino más bien sería esperable que ocurriese lo contrario, se detectaría una acumulación de dS. En contra de la acumulación constante de mutaciones, también está el resultado obtenido con los genomas completos, en el que se observaron diferentes tasas de evolución a lo largo de los pases de recuperación. Este resultado indicaba que el número de mutaciones acumuladas por pase había cambiado a lo largo de los pases de recuperación, aumentando a medida que se realizaban los pases seriados.

Por último, tras medir multitud de variables en estos virus, se realizó un análisis de regresión múltiple. En este análisis se obtuvo como variable más relacionada con el aumento de eficacia la heterogeneidad de las cuasiespecies junto con la tasa de evolución. Este resultado apoya la importancia de la heterogeneidad de las cuasiespecies como una de las principales fuerzas directrices de la recuperación de la eficacia biológica. Todos estos datos apuntaban a una recuperación de eficacia mediada por la generación de unas poblaciones cada vez más complejas. Esta complejidad por sí misma sería beneficiosa para los virus, como indicarían los incrementos de eficacia sin fijación de mutaciones en los primeros pases y, además, permitiría una mejor exploración del paisaje de secuencias con el objetivo de encontrar las mejores soluciones evolutivas.

Los resultados obtenidos en este estudio apoyan la teoría de las cuasiespecies para explicar la evolución viral y están de acuerdo con trabajos previos, en los que se ha demostrado una fuerte relación entre la variabilidad y la patogénesis viral. Otros trabajos en VIH-1 en pacientes han demostrado la relación entre aumento de complejidad y progresión de la enfermedad (Shankarappa y cols., 1999) o de la eficacia (Troyer y cols., 2005). Así mismo, los trabajos publicados por nuestro grupo anteriores a esta tesis también apoyan esta relación entre variabilidad y eficacia (Borderia y cols., 2010; Lorenzo-Redondo, Borderia y Lopez-Galindez, 2011). Resultados similares se encuentran en estudios con otros virus, como es el caso de VHC, en el que se ha visto una asociación positiva entre la variación en las zonas variables del genoma del virus y la progresión de la enfermedad en individuos infectados (Pawlotsky, 2006); incluso se ha observado una mayor susceptibilidad al tratamiento con IFN- α en poblaciones menos heterogéneas, mientras que ocurría lo contrario con la poblaciones más complejas (Pawlotsky, 2006). La importancia de la complejidad de las cuasiespecies en la patogénesis viral también ha sido demostrada en virus como el virus del Nilo Occidental (Ciota y cols., 2007) o en el virus de las paperas (Sauder y cols., 2006). Todos estos estudios demuestran que la complejidad de las cuasiespecies tiene un papel fundamental en las propiedades y evolución de las poblaciones virales, además de unas repercusiones clínicas y patogénicas importantes (Perales y cols., 2010).

5.3 Análisis de las fuerzas selectivas operando durante los pases de recuperación.

Tanto al estudiar las secuencias globales de los virus, como las cuasiespecies, se analizaron el tipo de mutaciones, sinónimas y no sinónimas, que aparecían, para poder conocer las fuerzas selectivas que podrían estar operando durante el proceso de recuperación de eficacia. Aunque existe cierta controversia sobre la utilización de las dN y dS para estudiar el tipo de selección, por ejemplo (Wong y cols., 2004), este tipo de análisis se utiliza de forma mayoritaria para examinar el tipo de selección detrás de la evolución observada.

Las mutaciones aparecidas en los genomas completos eran en su gran mayoría no sinónimas (80%). El porcentaje observado se podría considerar incluso mayor si se tienen en cuenta el gran número de mutaciones aparecidas en la LTR, que por el alto grado de estructuración de la región, podrían actuar también como mutaciones no sinónimas al afectar a dichas estructuras. Este resultado indica que, durante la recuperación de eficacia, estaría operando en gran medida la selección positiva. Esto se corroboró con los análisis de selección que mostraban una sobre-representación de mutaciones o una preponderancia de la selección positiva en todos los genes, al menos en uno de los pases analizados, excepto en *pol* (véase Figura 4.6C). Estos resultados eran de esperar, ya que el virus tiene que realizar cambios para recuperar la eficacia perdida, lo que tendría que ser obligatoriamente por medio de mutaciones no sinónimas. Estos cambios se produjeron a lo largo de todo el genoma, manteniendo el gen *pol* bajo selección negativa, debido a los efectos deletéreos que tienen los cambios en este gen.

Sin embargo, llamaba mucho más la atención el resultado obtenido en las cuasiespecies. Al analizar la composición de los espectros de mutantes, también se observó una dominancia de dN dentro de las cuasiespecies (Figura 4.10). Esta dominancia se incrementó a lo largo de los pases

de recuperación de forma casi exponencial. La gran mayoría de estas mutaciones no se imponían en la población, sino que eran variantes minoritarias de la cuasiespecie. Estos datos indicaban que la selección positiva también operaba a nivel de las cuasiespecies, moldeando la composición de las mismas. Así, los mutantes que componen la cuasiespecie están siendo también seleccionados, probablemente para generar un conjunto más favorable, aunque la mayoría de las mutaciones nunca lleguen a imponerse en la población viral.

Estos resultados indican que el proceso de recuperación de eficacia biológica “in vitro” del VIH-1 está dirigido principalmente por la selección positiva. El mecanismo seguido sería del tipo “darwiniano”, con generación de variabilidad global y selección de los variantes más eficaces, que continuarían generando más variabilidad. Este mecanismo ya había sido previamente descrito para el VIH-1 (Ganeshan y cols., 1997) y ha podido ser muy bien visualizado con la utilización de las redes filogenéticas mostradas en el apartado 4.3.5.

Dicho proceso, mantendría un balance entre la selección de mutaciones beneficiosas y la generación de variabilidad, que de no conservarse conduciría a pérdidas de eficacia. Los cuellos de botella demasiado severos producidos por mutaciones muy beneficiosas, casos de D1.5 y G1.5, llevarían a caídas de eficacia a largo plazo; mientras que una gran variabilidad sin encontrar los cambios oportunos, caso de K2, impediría al virus recuperar su eficacia biológica.

5.4 Fenotipo de los virus durante la recuperación de la eficacia biológica.

Al analizar cómo ha cambiado el fenotipo de los virus durante el proceso de recuperación biológica, cabe destacar el aumento de citopatogenicidad tan significativo observado a lo largo de los pases seriados. Esta elevada citopatogenicidad ha hecho disminuir considerablemente el tiempo de cultivo de los virus, lo que parece haber influido decisivamente en el resto de características virales observadas. Los virus, al producir antes la muerte celular de los cultivos, tienen menor tiempo para producir partículas virales, lo que explica las caídas observadas en los últimos 10 pases de los valores de título viral y de producción de p24 (Figura 4.23). Además, este aumento de citopatogenicidad puede estar ejerciendo una fuerza selectiva, que se vería reflejada en las mutaciones acumuladas en los últimos pases. De este modo, en los pases iniciales, se habrían acumulado mutaciones que incrementan la velocidad de replicación de los virus en los cultivos (mutaciones en *gag* o en *pol*). Por el contrario, parece que en los pases finales se habría favorecido la entrada de los virus al hospedador ante la menor cantidad de células disponibles debido al aumento de la citopatogenicidad de los virus (mutaciones en *env*, sobre todo gran aumento en gp41) (Figura 4.6). Así que las propias características de los virus en un momento dado, condicionarían la posterior evolución de los mismos.

Por otra parte, llaman la atención los resultados de la actividad RT, en los que se ha visto una caída de la actividad total a medida que aumentaban los pases, lo que resultaba inesperado (Figura 4.24). Al normalizar los valores por la producción de p24 el resultado era similar, lo que no ocurría cuando la normalización se llevaba a cabo con el título. Probablemente, estos resultados indicaban una alta especialización de los virus en los pases iniciales, que se ha ido perdiendo en favor de la generación de variabilidad a medida que se realizaban más pases. Esta hipótesis estaría en consonancia

con la caída de eficacia observada en los virus D1.5 y G1.5, en los que se observaban los valores más altos de actividad RT en el pase 21, indicando, junto a su elevado título, su gran adaptación en ese momento. Sin embargo, esta adaptación habría sido perjudicial al continuar los pases seriados. Por este motivo los virus habrían ido disminuyendo su actividad RT total, sin variar la calidad de la RT de los virus infecciosos, indicado por el mantenimiento de los valores de RT normalizada por el título. La caída vendría derivada de la mayor generación de virus con baja actividad RT al aumentar la heterogeneidad de la población.

Para finalizar, el examen del fenotipo viral nos ha mostrado un comportamiento similar de los parámetros estudiados en la mayoría de los virus: aumento de eficacia, aumento de citopatogenicidad, optimización de la producción de partículas virales, valores similares de título viral y caída de la actividad RT total. Esto estaría indicando que los diferentes caminos evolutivos seguidos por los distintos virus, han conducido a efectos fenotípicos similares. Resultado en concordancia con otros trabajos de evolución en virus ARN, en los que se ha visto que diferentes mutaciones pueden estar asociadas con el mismo comportamiento fenotípico (Domingo y Holland, 1997).

5.5 Mutaciones aparecidas durante la pérdida y recuperación de eficacia “in vitro” del VIH-1 estudiadas por mutagénesis dirigida.

En este estudio, se seleccionaron varias mutaciones aparecidas durante el proceso de pérdida de eficacia, debida a los pases placa a placa, y en la posterior recuperación de eficacia por pases seriados en gran tamaño poblacional, que parecían tener importancia en todo el proceso. Se eligieron varias mutaciones aparecidas durante el Trinquete de Muller, otras fijadas al realizar las semillas de los clones tras los pases placa a placa y, por último, cambios aparecidos durante los pases de recuperación. Las mutaciones se eligieron por diversos motivos. Las mutaciones en LTR G379A y A586G, por estar en regiones importantes para la replicación viral (sitio de unión a factores de transcripción y sitio de unión al cebador, respectivamente). Los cambios en la RT se eligieron por su proximidad al centro catalítico y su probable repercusión en la replicación viral. Otras mutaciones se seleccionaron por ser comunes a varios virus, como la A587G (LTR), la G593T (LTR), la V35I (p17) y la S314K (gp120), además de tener efectos importantes descritos anteriormente.

Los resultados mostrados en la Figura 4.32 indican, como era de esperar, que las mutaciones aparecidas durante los cuellos de botella poblacionales tienen un alto coste de eficacia, disminuyendo mucho la infectividad de los virus. Las mutaciones elegidas se situaban todas en la RT, excepto una en la región de unión a los promotores en la LTR. La mera adición de una de estas mutaciones significaba un perjuicio importante para el virus, por lo que podrían ser muy buenas candidatas para incluir en virus con el fin de obtener variantes atenuadas.

Por su parte, todas las mutaciones elegidas de las aparecidas en las semillas se situaban en la LTR. La acumulación de mutaciones en esta región parecía ser un recurso habitual de los virus para una rápida recuperación inicial de eficacia, ya que todos los virus analizados presentaban alguna mutación en esta región al hacer las semillas. Estos cambios producen una mejora moderada de la infectividad, pero al aparecer en la LTR deben de suponer una forma más rápida de aumentar la

eficacia, debido a la acción directa de mutaciones en esta zona sobre la replicación. En el caso de la mutación A586G, además se producía un aumento en la citopatogenicidad del virus, lo que sería beneficioso para el virus en los primeros pasos de la recuperación. Estas mejoras permiten a los virus generar una variabilidad inicial, a partir de la cuál continuar la recuperación.

Finalmente, el análisis de las mutaciones fijadas a lo largo de los pases de recuperación mostró que producían un aumento muy significativo en la infectividad y la citopatogenicidad de los virus que las portaban. Sin embargo, estos cambios no producían aumentos estadísticamente significativos en la eficacia biológica con respecto al virus control. Además, excepto en el caso del mutante S314K, los valores obtenidos eran muy similares al control s61. Especial interés merece la mutación V35I de la p17, que aparecía en virus de 3 linajes diferentes, ya que es punto común en la recuperación de varios virus. Esta mutación parece haber sido un recurso habitual en la recuperación de los cuellos de botella poblacionales sufridos por los virus estudiados, ya que produce un gran aumento en el título viral, lo que permitiría generar la diversidad perdida por la extremada reducción de la población que conllevan dichos cuellos de botella. Esta mutación ya había sido descrita en otros trabajos “*in vitro*” como acompañante de mutaciones de resistencia a la proteasa para mejorar su eficacia (Gatanaga y cols., 2002), indicando su importancia en la recuperación de eficacia. Sin embargo, las bases moleculares de los efectos producidos por este cambio no han sido dilucidadas en el presente estudio, ya que se necesitaría un análisis mucho más exhaustivo que se escapa a los propósitos de esta tesis. De todas formas, se pueden proponer dos posibles hipótesis:

En primer lugar, al localizarse este cambio en el motivo NLS de transporte al núcleo de la proteína p17, un cambio en este motivo podría suponer una mejora en el reconocimiento de esta secuencia y producir un transporte más eficaz, lo que aumentaría la velocidad del ciclo viral y, por lo tanto, un aumento en la producción de virus.

La segunda hipótesis tendría que ver con la unión a la membrana de la proteína p17. Si este cambio, que se encuentra en la parte exterior del complejo proteico de la p17, mejorase dicha unión, se podría mejorar el proceso de gemación, lo que también aumentaría la producción de partículas virales. De cualquier forma, es una mutación que requiere más estudio, ya que puede ser clave para entender mejor los mecanismos seguidos por el VIH-1 para la recuperación de eficacia.

En conclusión, estos análisis de mutaciones nos han permitido desarrollar un mejor conocimiento del proceso de pérdida y recuperación de eficacia biológica debidas a cambios en el tamaño poblacional. La drástica disminución de tamaño poblacional derivada de cuellos de botella, permite la acumulación de mutaciones deletéreas, cuya simple presencia conduce a perjuicios importantes en el virus. Los virus que incorporaron estas mutaciones durante el Trinquete de Muller, tuvieron recuperaciones más tardías que el resto (virus de los linajes I y K), por la dificultad de revertir sus efectos. Esto demuestra que los virus son muy sensibles a ciertas mutaciones, siendo lógico que varias de estas mutaciones las hayamos encontrado en la RT.

Por su parte, la recuperación de eficacia biológica se ha producido por medio de mutaciones que mejoraban la producción y el título viral, pero que no producían aumentos significativos en la eficacia. Esto apoya nuestra hipótesis de la gran importancia de la heterogeneidad de las cuasiespecies en la recuperación. La gran mayoría de las mutaciones únicas no producen cambios en la eficacia, sino que se fijan por producir mayor cantidad de virus para generar una mayor complejidad de la

población, factor que lleva a la recuperación de eficacia. Este resultado ha de ser tenido muy en cuenta en posibles tratamientos o vacunas, ya que se puede disminuir la eficacia con mutaciones puntuales, pero el virus es capaz de recuperarse de múltiples maneras, siendo fundamental la generación de diversidad. Por ello, el simple análisis de secuencias consenso en los pacientes, sin tener en cuenta el espectro de mutantes del que está compuesto la población, puede llevar a malos diagnósticos y fallos en la terapia. Además, para la generación de vacunas hay que tener en cuenta esta inmensa variabilidad, lo que ya está ocurriendo en los últimos intentos de generación de vacunas eficaces contra el VIH-1, incluyendo diversos epítomos en las mismas (Gaschen y cols., 2002).

5.6 Generación del paisaje de eficacias experimental del proceso de recuperación de eficacia “in vitro” del VIH-1.

En la parte final de este estudio y debido al conocimiento tan completo que se tenía de los virus analizados, se investigó la posibilidad de desarrollar una metodología para la generación de un paisaje de eficacias que describiese la evolución seguida por los virus en su recuperación de eficacia biológica. La primera aproximación llevada a cabo fue utilizando métodos filogenéticos para la separación de las distintas secuencias, aplicando luego la eficacia calculada de cada una como tercera dimensión. La generación del paisaje de eficacias utilizando esta metodología permitía una atractiva visualización de la evolución de los virus estudiados. El paisaje generado era rugoso, en concordancia con trabajos previos tanto teóricos (Elena y Sanjuan, 2003; Macken y Perelson, 1989) como los basados en datos experimentales (Burch y Chao, 1999; Hayashi y cols., 2006), conteniendo diferentes picos y valles. Además, el pico generado en la zona del virus original, que correspondería al fenotipo salvaje, estaba alejado del resto de los picos y ejercía como atractor de las secuencias en su entorno (véase en las Figuras 4.33 y 4.34 como retornan los virus del linaje I al pico original a diferencia del resto de los virus) resultado similar al de otros trabajos (Pitt y Ferre-D’Amare, 2010). Con este paisaje podíamos seguir la evolución de los virus y describir visualmente el proceso de recuperación. Además, era deseable que estos paisajes tuvieran la capacidad de estimación de eficacias desconocidas como posible futura herramienta de trabajo. Sin embargo, este tipo de paisajes no nos permitía estimar eficacias desconocidas, ya que la utilización de métodos filogenéticos, como se explicó en el apartado 4.7.1, no genera topologías fijas, lo que impide su utilización como método predictivo.

Con este fin, se llevó a cabo una aproximación diferente haciendo uso de redes neuronales artificiales, que ya habían sido ampliamente utilizadas para la clasificación de todo tipo de datos. Utilizando un tipo de redes auto-organizativas, las SOM, pudimos generar un paisaje de eficacias, que ya permitía la estimación de eficacias desconocidas al distribuir las secuencias en la red de tamaño establecido que utilizan (en nuestro caso de 15 x 15 neuronas), por lo tanto de topología fija. Este nuevo paisaje era muy similar al generado por filogenia, lo que indicaba la robustez de ambas aproximaciones.

Al utilizar esta metodología para estimar las eficacias de cada variante de las cuasiespecies virales, se pudo observar una gran concordancia entre la estimación por SOM y lo observado en los pases “in vitro”. Los mutantes con mayor eficacia estimada, acababan imponiéndose en la población,

mientras que soluciones menos eficaces son abandonadas (Figura 4.35C). En el seguimiento de la evolución de las cuasiespecies que permiten establecer estos paisajes, se puede observar la generación de una gran diversidad para encontrar las mutaciones más beneficiosas y el desplazamiento posterior de la mayor parte de la población a estos puntos del paisaje con mayor eficacia. En dichas cuasiespecies se puede observar la presencia de variantes minoritarios correspondientes a los virus iniciales constituyendo la memoria de la población, como ya se había descrito para las cuasiespecies virales (Briones y cols., 2006; Ruiz-Jarabo y cols., 2000).

El desarrollo de este método de estimación de eficacias puede resultar de gran utilidad, debido a que el tiempo y exigencia computacional son pequeños y su error estimado (en torno al 11%) comparable al experimental. De esta forma, se podría calcular experimentalmente la eficacia asociada a diversas secuencias y estimar las eficacias desconocidas de secuencias relacionadas. Con el uso cada vez más extendido de la secuenciación masiva, esta herramienta de estimación de eficacias podría llegar a ser de gran utilidad, ya que el cálculo experimental de todos los miembros de la cuasiespecies es completamente inviable. Así, se podría calcular la eficacia de algunos genomas y luego estimar el resto del espectro de mutantes, siempre teniendo en cuenta la impredecibilidad de los virus ARN, en los que un único cambio puede producir efectos extremos. Con la ayuda de estos paisajes generados, se podrían predecir fallos en el tratamiento de pacientes, cambios de fenotipo o estudiar la evolución de cualquier sistema biológico que tenga propiedades asociadas directamente a su secuencia nucleotídica.

5.7 Discusión general.

Tras finalizar todos los análisis llevados a cabo para completar la presente tesis, hemos podido concluir que la recuperación de eficacia “in vitro” del VIH-1 es un proceso complejo y dinámico, en el que los diferentes virus han seguido múltiples caminos evolutivos para alcanzar la mencionada recuperación. Durante esta recuperación, el patrón común es el de generación de complejidad poblacional para aumentar la eficacia biológica y selección de los variantes más eficaces. El proceso sigue una dinámica “darwiniana” dominada en todo momento por la selección positiva, que opera tanto a nivel de secuencias consenso como de cuasiespecies. Las propias características del virus parecen ser la principal fuente de selección junto con la competición entre variantes dentro de la cuasiespecie, que se detecta por la alternancia de nucleótidos en la misma posición en varios puntos del genoma a lo largo de los pases de recuperación. Todos los resultados obtenidos confirman el importante papel de las cuasiespecies en la evolución y en las características del VIH-1, sugiriendo un papel principal de las mismas en la recuperación de eficacia biológica tras disminuciones drásticas del tamaño poblacional. Por lo tanto, entender el comportamiento de este espectro de mutantes generado en las poblaciones virales parece fundamental a la hora de comprender el transcurso de la infección por el virus y los posibles mecanismos para lograr combatirlo. Además, esta tesis apoya experimentalmente la validez del modelo de cuasiespecies para explicar la evolución de las poblaciones virales y las características desplegadas por los virus durante la infección del hospedador.

6. Conclusiones

- Pases “in vitro” en gran tamaño poblacional producen la recuperación de la eficacia biológica de clones del VIH-1 debilitados por pases en cuello de botella. Las características de esta recuperación son muy diversas entre los virus estudiados.
- El proceso de recuperación de eficacia está dominado por la selección positiva, que condiciona la fijación de mutaciones en los genomas globales y la composición de las cuasiespecies.
- Las mutaciones que aparecen en los primeros 10 pases de recuperación son fijadas mayoritariamente en cada virus y constituyen la base a partir de la cual se genera una mayor complejidad poblacional, fijándose como consecuencia nuevas mutaciones.
- A pesar de la diversidad de mutaciones aparecidas en los diferentes virus, se ha observado la aparición de 9 mutaciones convergentes, algunas de ellas asociadas con importantes efectos fenotípicos. Entre ellas, se encontró la mutación V35I de la proteína p17, que apareció en 6 virus de tres linajes diferentes.
- La acumulación de deleciones en la región variable V1-V2 de la envuelta es un mecanismo común de recuperación de eficacia biológica en estos virus.
- La tasa de evolución y la heterogeneidad de las cuasiespecies son las variables más directamente relacionadas con el aumento de eficacia biológica.
- Los pases seriados de recuperación conducen a un aumento muy significativo de la citopatogenicidad de los virus estudiados.
- Aunque globalmente la eficacia sigue creciendo tras 30 pases de recuperación, excepcionalmente, dos virus con características de gran adaptación en el pase 21, han perdido eficacia de manera significativa en los últimos 10 pases (pase 31).
- La utilización de métodos filogenéticos y de mapas auto-organizativos ha permitido generar paisajes de eficacias del proceso de recuperación de eficacia “in vitro” del VIH-1. Los paisajes obtenidos con los mencionados mapas poseen capacidad de estimación de eficacias desconocidas.
- La heterogeneidad de las cuasiespecies virales juega un papel fundamental en la recuperación de la eficacia biológica de virus de VIH-1 previamente debilitados, apoyando de este modo la validez del modelo de cuasiespecies en poblaciones virales.

7. Bibliografía.

- Altschul, S. F., Gish, W., Miller, W., Myers, E. W. y Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol* **215**(3), 403-10.
- Anastassopoulou, C. G. y Kostrikis, L. G. (2005). Viral correlates of HIV-1 disease. *Curr HIV Res* **3**(2), 113-32.
- Anisimova, M. y Gascuel, O. (2006). Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Syst Biol* **55**(4), 539-52.
- Baltimore, D. (1970). RNA-dependent DNA polymerase in virions of RNA tumour viruses. *Nature* **226**(5252), 1209-11.
- Balzarini, J., Pelemans, H., Riess, G., Roesner, M., Winkler, I., De Clercq, E. y Kleim, J. P. (1997). Zidovudine-resistant human immunodeficiency virus type 1 strains subcultured in the presence of both lamivudine and quinoxaline HBY 097 retain marked sensitivity to HBY 097 but not to lamivudine. *J Infect Dis* **176**(5), 1392-7.
- Barre-Sinoussi, F., Chermann, J. C., Rey, F., Nugeyre, M. T., Chamaret, S., Gruest, J., Dauguet, C., Axler-Blin, C., Vezinet-Brun, F., Rouzioux, C., Rozenbaum, W. y Montagnier, L. (1983). Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science* **220**(4599), 868-71.
- Beerens, N., Groot, F. y Berkhout, B. (2001). Initiation of HIV-1 reverse transcription is regulated by a primer activation signal. *J Biol Chem* **276**(33), 31247-56.
- Berkowitz, R., Fisher, J. y Goff, S. P. (1996). RNA packaging. *Curr Top Microbiol Immunol* **214**, 177-218.
- Biebricher, C. K. y Eigen, M. (2005). The error threshold. *Virus Res* **107**(2), 117-27.
- Borderia, A. V., Lorenzo-Redondo, R., Pernas, M., Casado, C., Alvaro, T., Domingo, E. y Lopez-Galindez, C. (2010). Initial fitness recovery of HIV-1 is associated with quasispecies heterogeneity and can occur without modifications in the consensus sequence. *PLoS One* **5**(4), e10319.
- Bour, S., Schubert, U. y Strebel, K. (1995). The human immunodeficiency virus type 1 Vpu protein specifically binds to the cytoplasmic domain of CD4: implications for the mechanism of degradation. *J Virol* **69**(3), 1510-20.
- Briones, C., de Vicente, A., Molina-Paris, C. y Domingo, E. (2006). Minority memory genomes can influence the evolution of HIV-1 quasispecies in vivo. *Gene* **384**, 129-38.
- Brodsky, L. I., Vasilyev, A. V., Kalaydzidis, Y. L., Osipov, Y. S., Tatuzov, A. R. L. y Feranchuk, S. I. (1992). GeneBee: the Program Package for Biopolymer Structure Analysis. *Dimacs* **8**, 127-139.
- Brown, P. O., Bowerman, B., Varmus, H. E. y Bishop, J. M. (1989). Retroviral integration: structure of the initial covalent product and its precursor, and a role for the viral IN protein. *Proc Natl Acad Sci U S A* **86**(8), 2525-9.
- Bruggeman, J., Debets, A. J., Wijngaarden, P. J., deVisser, J. A. y Hoekstra, R. F. (2003). Sex slows down the accumulation of deleterious mutations in the homothallic fungus *Aspergillus nidulans*. *Genetics* **164**(2), 479-85.
- Bukrinsky, M. I., Haggerty, S., Dempsey, M. P., Sharova, N., Adzhubel, A., Spitz, L., Lewis, P., Goldfarb, D., Emerman, M. y Stevenson, M. (1993). A nuclear localization signal within HIV-1 matrix protein that governs infection of non-dividing cells. *Nature* **365**(6447), 666-9.
- Burch, C. L. y Chao, L. (1999). Evolution by small steps and rugged landscapes in the RNA virus phi6. *Genetics* **151**(3), 921-7.
- Burch, C. L. y Chao, L. (2000). Evolvability of an RNA virus is determined by its mutational neighbourhood. *Nature* **406**(6796), 625-8.
- Chao, L. (1990). Fitness of RNA virus decreased by Muller's ratchet. *Nature (London)* **348**(6300), 454-5.
- Chen, M. Y., Maldarelli, F., Karczewski, M. K., Willey, R. L. y Strebel, K. (1993). Human immunodeficiency virus type 1 Vpu protein induces degradation of CD4 in vitro: the cytoplasmic domain of CD4 contributes to Vpu sensitivity. *J Virol* **67**(7), 3877-84.

- Choe, H., Farzan, M., Sun, Y., Sullivan, N., Rollins, B., Ponath, P. D., Wu, L., Mackay, C. R., LaRosa, G., Newman, W., Gerard, N., Gerard, C. y Sodroski, J. (1996).** The beta-Chemokine Receptors CCR3 and CCR5 Facilitate Infection by Primary HIV-1 Isolates. *Cell* **85**, 1135-1148.
- Chowers, M. Y., Spina, C. A., Kwoh, T. J., Fitch, N. J. S., Richman, D. D. y Guatelli, J. C. (1994).** Optimal Infectivity in vitro of Human Immunodeficiency virus type 1 requires an intact nef gene. *Journal of Virology* **68**, 2906-2914.
- Ciota, A. T., Ngo, K. A., Lovelace, A. O., Payne, A. F., Zhou, Y., Shi, P. Y. y Kramer, L. D. (2007).** Role of the mutant spectrum in adaptation and replication of West Nile virus. *J Gen Virol* **88**(Pt 3), 865-74.
- Clarke, D. K., Duarte, E. A., Moya, A., Elena, S. F., Domingo, E. y Holland, J. (1993).** Genetic bottlenecks and population passages cause profound fitness differences in RNA viruses. *Journal of Virology* **67**(1), 222-228.
- Crawford, S. y Goff, S. P. (1985).** A deletion mutation in the 5' part of the pol gene of Moloney murine leukemia virus blocks proteolytic processing of the gag and pol polyproteins. *J Virol* **53**(3), 899-907.
- Dayton, A. I., Sodroski, J. G., Rosen, C. A., Goh, W. C. y Haseltine, W. A. (1986).** The trans-activator gene of the human T cell lymphotropic virus type III is required for replication. *Cell* **44**(6), 941-7.
- De Jong, J. J., De Ronde, A., Keulen, W., Tersmette, M. y Goudsmit, J. (1992).** Minimal requirements for the human immunodeficiency virus type 1 V3 domain to support the syncytium-inducing phenotype: analysis by single amino acid substitution. *J Virol* **66**(11), 6777-80.
- de la Torre, J. C. y Holland, J. J. (1990).** RNA virus quasispecies populations can suppress vastly superior mutant progeny. *J. Virol* **64**, 6278-6281.
- Deforche, K., Camacho, R., Van Laethem, K., Lemey, P., Rambaut, A., Moreau, Y. y Vandamme, A. M. (2008).** Estimation of an in vivo fitness landscape experienced by HIV-1 under drug selective pressure useful for prediction of drug resistance evolution during treatment. *Bioinformatics* **24**(1), 34-41.
- Delgado, S. (2009).** Nuevos algoritmos de entrenamiento y visualización de redes neuronales autoorganizativas para su aplicación en Teledetección. *Tesis Doctoral*.
- Deng, H., Liu, R., Ellmeier, W., Choe, S., Unutmaz, D., Burkhart, M., Di Marzio, P., Marmon, S., Sutton, R. E., Hill, C. M., Davis, C. B., Peiper, S. C., Schall, T. J., Littman, D. R. y Landau, N. R. (1996).** Identification of a major co-receptor for primary isolates of HIV-1. *Nature* **381**, 661-666.
- Derdeyn, C. A., Decker, J. M., Bibollet-Ruche, F., Mokili, J. L., Muldoon, M., Denham, S. A., Heil, M. L., Kasolo, F., Musonda, R., Hahn, B. H., Shaw, G. M., Korber, B. T., Allen, S. y Hunter, E. (2004).** Envelope-constrained neutralization-sensitive HIV-1 after heterosexual transmission. *Science* **303**(5666), 2019-22.
- Diaz Arenas, C. y Lehman, N.** Quasispecies-like behavior observed in catalytic RNA populations evolving in a test tube. *BMC Evol Biol* **10**, 80.
- Dobzhansky, T. (1964).** Biology, Molecular and Organismic. *Am Zool* **4**, 443-52.
- Domingo, E. (2002).** Quasispecies Theory in Virology. *Journal of Virology* **76**(1), 463-465.
- Domingo, E., Escarmis, C., Lazaro, E. y Manrubia, S. C. (2005).** Quasispecies dynamics and RNA virus extinction. *Virus Research* **107**(2), 129-139.
- Domingo, E., Escarmis, C., Martínez, M. A., Martínez-Salas, E. y Mateu, M. G. (1992).** Foot-and-mouth disease virus populations are quasispecies. *Curr Top Microbiol Immunol* **176**, 33-47.
- Domingo, E., Escarmis, C., Menendez-Arias, L. y Holland, J. J. (1999).** Viral Quasispecies and Fitness Variations. En "Origin and Evolution of Viruses" (E. Domingo, R. Webster, and J. Holland, Eds.), pp. 141-161. Academic Press.

- Domingo, E. y Holland, J. J.** (1997). RNA Virus Mutations and Fitness for Survival. *Annual Review of Microbiology* **51**, 151-178.
- Domingo, E., Martinez-Salas, E., Sobrino, F., de la Torre, J. C., Portela, A., Ortin, J., Lopez-Galindez, C., Perez-Brena, P., Villanueva, N., Najera, R. y et al.** (1985). The quasispecies (extremely heterogeneous) nature of viral RNA genome populations: biological relevance--a review. *Gene* **40**(1), 1-8.
- Domingo, E., Sabo, D., Taniguchi, T. y Weissmann, C.** (1978). Nucleotide sequence heterogeneity of an RNA phage population. *Cell* **13**(4), 735-44.
- Doranz, B. J., Rucker, J., Yi, Y., Smyth, R. J., Samson, M., Peiper, S. C., Parmentier, M., Collman, R. G. y Doms, R. W.** (1996). A dual-tropic primary HIV-1 isolate that uses fusin and the beta-chemokine receptors CKR-5, CKR-3, and CKR-2b as fusion cofactors. *Cell* **85**(7), 1149-58.
- Drummond, A. J. y Rambaut, A.** (2007). BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* **7**, 214.
- Duarte, E., Clarke, D., Moya, A., Domingo, E. y Holland, J. J.** (1992). Rapid fitness losses in mamalian RNA virus due Muller's ratchet. *Proc.Natl.Acad. Sci.USA* **89**, 6015-6019.
- DuBridge, R. B., Tang, P., Hsia, H. C., Leong, P. M., Miller, J. H. y Calos, M. P.** (1987). Analysis of mutation in human cells by using an Epstein-Barr virus shuttle system. *Mol Cell Biol* **7**(1), 379-87.
- Edgar, R. C.** (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**(5), 1792-7.
- Eigen, M.** (1971). Selforganization of matter and the evolution of biological macromolecules. *Naturwissenschaften* **58**(10), 465-523.
- Eigen, M.** (1996). On the nature of virus quasispecies [letter; comment]. *Trends Microbiol* **4**(6), 216-8.
- Eigen, M. y Biebricher, C. K.** (1988). "Sequence space and quasispecies distribution." RNA genetics (E. Domingo, J. J. Holland, and P. Ahlquist, Eds.), III, Variability of RNA Genomes. CRC Press, Boca Raton, Florida.
- Elena, S. F. y Sanjuan, R.** (2003). Evolution. Climb every mountain? *Science* **302**(5653), 2074-5.
- Escarmis, C., Davila, M., Charpentier, N., Bracho, A., Moya, A. y Domingo, E.** (1996). Genetic lesions associated with Muller's Ratchet in an RNA virus. *J Mol Biol* **264**, 255-67.
- Escarmis, C., Davila, M. y Domingo, E.** (1999). Multiple molecular pathways for fitness recovery of an RNA virus debilitated by operation of Muller's ratchet. *J Mol Biol* **285**(2), 495-505.
- Escarmis, C., Lazaro, E., Arias, A. y Domingo, E.** (2008). Repeated bottleneck transfers can lead to non-cytocidal forms of a cytopathic virus: implications for viral extinction. *J Mol Biol* **376**(2), 367-79.
- Essajee, S. M., Pollack, H., Rochford, G., Oransky, I., Krasinski, K. y Borkowsky, W.** (2000). Early changes in quasispecies repertoire in HIV-infected infants: correlation with disease progression. *AIDS Res Hum Retroviruses* **16**(18), 1949-57.
- Feng, X., Li, Y., Gu, J., Zhuo, Y. y Yang, H.** (2007). Error thresholds for quasispecies on single peak Gaussian-distributed fitness landscapes. *J Theor Biol* **246**(1), 28-32.
- Feng, Y., Broder, C. C., Kennedy, P. E. y Berger, E. A.** (1996). HIV-1 Entry Cofactor: Functional cDNA Cloning of a Seven-Transmembrane, G Protein-Coupled Receptor. *Science* **272**, 872-877.
- Fernandez, G., Clotet, B. y Martinez, M. A.** (2007). Fitness landscape of human immunodeficiency virus type 1 protease quasispecies. *J Virol* **81**(5), 2485-96.
- Fisher, A. G., Feinberg, M. B., Josephs, S. F., Harper, M. E., Marselle, L. M., Reyes, G., Gonda, M. A., Aldovini, A., Debouk, C., Gallo, R. C. y et al.** (1986). The trans-activator gene of HTLV-III is essential for virus replication. *Nature* **320**(6060), 367-71.
- Freed, E. O., Englund, G. y Martin, M. A.** (1995). Role of the basic domain of human immunodeficiency virus type 1 matrix in macrophage infection. *J Virol* **69**(6), 3949-54.

- Fujiwara, T. y Mizuuchi, K. (1988). Retroviral DNA integration: structure of an integration intermediate. *Cell* **54**(4), 497-504.
- Gallo, R. C., Salahuddin, S. Z., Popovic, M., Shearer, G. M., Kaplan, M., Haynes, B. F., Palker, T. J., Redfield, R., Oleske, J., Safai, B. y et al. (1984). Frequent detection and isolation of cytopathic retroviruses (HTLV-III) from patients with AIDS and at risk for AIDS. *Science* **224**(4648), 500-3.
- Ganeshan, S., Dickover, R. E., Korber, B. T., Bryson, Y. J. y Wolinsky, S. M. (1997). Human immunodeficiency virus type 1 genetic evolution in children with different rates of development of disease. *J Virol* **71**(1), 663-77.
- Garcia, J. V. y Miller, A. D. (1991). Serine phosphorylation-independent downregulation of cell-surface CD4 by nef. *Nature* **350**(6318), 508-11.
- Gaschen, B., Taylor, J., Yusim, K., Foley, B., Gao, F., Lang, D., Novitsky, V., Haynes, B., Hahn, B. H., Bhattacharya, T. y Korber, B. (2002). Diversity considerations in HIV-1 vaccine selection. *Science* **296**(5577), 2354-60.
- Gatanaga, H., Suzuki, Y., Tsang, H., Yoshimura, K., Kavlick, M. F., Nagashima, K., Gorelick, R. J., Mardy, S., Tang, C., Summers, M. F. y Mitsuya, H. (2002). Amino acid substitutions in Gag protein at non-cleavage sites are indispensable for the development of a high multitude of HIV-1 resistance against protease inhibitors. *J Biol Chem* **277**(8), 5952-61.
- Gilbert, M. T. P., Rambaut, A., Wlasiuk, G., Spira, T. J., Pitchenik, A. E. y Worobey, M. (2007). The emergence of HIV/AIDS in the Americas and beyond. *Proc Natl Acad Sci U S A* **104**(47), 18566-18570
- Gonzalez-Garcia, I., Sole, R. V. y Costa, J. (2002). Metapopulation dynamics and spatial heterogeneity in cancer. *Proc Natl Acad Sci U S A* **99**(20), 13085-9.
- Gottlieb, M. S., Schroff, R., Schanker, H. M., Weisman, J. D., Fan, P. T., Wolf, R. A. y Saxon, A. (1981). Pneumocystis carinii pneumonia and mucosal candidiasis in previously healthy homosexual men: evidence of a new acquired cellular immunodeficiency. *N Engl J Med* **305**(24), 1425-31.
- Göttliger, H. (2001). HIV-1 Gag: a Molecular Machine Driving Viral Particle Assembly and Release. *HIV Sequence Compendium 2001.*, 2-28.
- Göttliger, H. G., Dorfman, T., Sodroski, J. G. y Haseltine, W. A. (1991). Effect of mutations affecting the p6 gag protein on human immunodeficiency virus particle release. *Proc Natl Acad Sci U S A* **88**(8), 3195-9.
- Guex, N. y Peitsch, M. C. (1997). SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* **18**(15), 2714-23.
- Guindon, S. y Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* **52**(5), 696-704.
- Harada, S., Koyanagi, Y. y Yamamoto, N. (1985). Infection of HTLV-III/LAV in HTLV-I-carrying cells MT-2 and MT-4 and application in a plaque assay. *Science* **229**, 563-6.
- Hayashi, Y., Aita, T., Toyota, H., Husimi, Y., Urabe, I. y Yomo, T. (2006). Experimental rugged fitness landscape in protein sequence space. *PLoS One* **1**, e96.
- Heeney, J. L., Dalgleish, A. G. y Weiss, R. A. (2006). Origins of HIV and the evolution of resistance to AIDS. *Science* **313**(5786), 462-6.
- Heinzinger, N. K., Bukinsky, M. I., Haggerty, S. A., Ragland, A. M., Kewalramani, V., Lee, M. A., Gendelman, H. E., Ratner, L., Stevenson, M. y Emerman, M. (1994). The Vpr protein of human immunodeficiency virus type 1 influences nuclear localization of viral nucleic acids in nondividing host cells. *Proc Natl Acad Sci U S A* **91**(15), 7311-5.
- Hinkley, T., Martins, J., Chappey, C., Haddad, M., Stawiski, E., Whitcomb, J. M., Petropoulos, C. J. y Bonhoeffer, S. (2011). A systems analysis of mutational effects in HIV-1 protease and reverse transcriptase. *Nat Genet* **43**(5), 487-9.
- Hofacker, I. L. (2003). Vienna RNA secondary structure server. *Nucleic Acids Res* **31**(13), 3429-31.

- Holland, J. J., de la Torre, J. C., Clarke, D. K. y Duarte, E. (1991). Quantitation of relative fitness and great adaptability of clonal populations of RNA viruses. *J Virol* **65** (6), 2960-7.
- Holland, J. J., Spindler, K., Horodyski, F., Gradau, E., Nichol, S. y Vande Pol, S. (1982). Rapid evolution of RNA genomes. *Science* **215**, 1577-1585.
- Holmes, E. C. (2010). The RNA virus quasispecies: fact or fiction? *J Mol Biol* **400**(3), 271-3.
- Holmes, E. C. y Moya, A. (2002). Is the quasispecies concept relevant to RNA viruses? *J Virol* **76**(1), 460-5.
- Hu, W. S. y Temin, H. M. (1990). Genetic consequences of packaging two RNA genomes in one retroviral particle: pseudodiploidy and high rate of genetic recombination. *Proc Natl Acad Sci U S A* **87**(4), 1556-60.
- Huang, M., Orenstein, J. M., Martin, M. A. y Freed, E. O. (1995). p6Gag is required for particle production from full-length human immunodeficiency virus type 1 molecular clones expressing protease. *J Virol* **69**(11), 6810-8.
- Hubert, J. J. (1984). Bioassay (D. Hunt, Ed.), pp. 65-66.
- Huson, D. H. y Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol* **23**(2), 254-67.
- Iglesias-Sanchez, M. J. y Lopez-Galindez, C. (2002). Analysis, Quantification and Evolutionary Consequences of the HIV-1 in vitro Recombination. *Virology* **304**, 392-402.
- Jacobo-Molina, A., Ding, J., Nanni, R. G., Clark, A. D., Jr., Lu, X., Tantillo, C., Williams, R. L., Kamer, G., Ferris, A. L., Clark, P., Hizi, A., Hughes, S. H. y Arnold, E. (1993). Crystal structure of human immunodeficiency virus type 1 reverse transcriptase complexed with double-stranded DNA at 3.0 Å resolution shows bent DNA. *Proc Natl Acad Sci U S A* **90**(13), 6320-4.
- Jung, A., Maier, R., Vartanian, J. P., Bocharov, G., Jung, V., Fischer, U., Meese, E., Wain-Hobson, S. y Meyerhans, A. (2002). Recombination: Multiply infected spleen cells in HIV patients. *Nature* **418**(6894), 144.
- Katoh, I., Yoshinaka, Y., Rein, A., Shibuya, M., Odaka, T. y Oroszlan, S. (1985). Murine leukemia virus maturation: protease region required for conversion from "immature" to "mature" core form and for virus infectivity. *Virology* **145**(2), 280-92.
- Keightley, P. D. y Caballero, A. (1997). Genomic mutation rates for lifetime reproductive output and lifespan in *Caenorhabditis elegans*. *Proc Natl Acad Sci U S A* **94**(8), 3823-7.
- Kibota, T. T. y Lynch, M. (1996). Estimate of the genomic mutation rate deleterious to overall fitness in *E. coli*. *Nature* **381**, 694-696.
- Kimura, M. (1989). The neutral theory of molecular evolution and the world view of the neutralists. *Genome* **31**(1), 24-31.
- Kleim, J. P., Rosner, M., Winkler, I., Paessens, A., Kirsch, R., Hsiou, Y., Arnold, E. y Riess, G. (1996). Selective pressure of a quinoxaline nonnucleoside inhibitor of human immunodeficiency virus type 1 (HIV-1) reverse transcriptase (RT) on HIV-1 replication results in the emergence of nucleoside RT-inhibitor-specific (RT Leu-74-->Val or Ile and Val-75-->Leu or Ile) HIV-1 mutants. *Proc Natl Acad Sci U S A* **93**(1), 34-8.
- Kohonen, T. (2001). "Self-Organizing Maps" (T. Kohonen, Ed.), (3rd edition): .
- Korber, B., Gaschen, B., Yusim, K., Thakallapally, R., Kesmir, C. y Detours, V. (2001). Evolutionary and immunological implications of contemporary HIV-1 variation. *Br Med Bull* **58**, 19-42.
- Korber, B., Muldoon, M., Theiler, J., Gao, F., Gupta, R., Lapedes, A., Hahn, B. H., Wolinsky, S. y Bhattacharya, T. (2000). Timing the ancestor of the HIV-1 pandemic strains [see comments]. *Science* **288**(5472), 1789-96.
- Kremer, M. y Schnierle, B. S. (2005). HIV-1 Vif: HIV's weapon against the cellular defense factor APOBEC3G. *Curr HIV Res* **3**(4), 339-44.
- Kuroda, M., Katano, H., Nakajima, N., Tobiume, M., Ainai, A., Sekizuka, T., Hasegawa, H., Tashiro, M., Sasaki, Y., Arakawa, Y., Hata, S., Watanabe, M. y Sata, T. (2009). Characterization of quasispecies of pandemic 2009 influenza A virus (A/H1N1/2009) by de

- novo sequencing using a next-generation DNA sequencer. *PLoS One* **5**(4), e10256.
- Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J. y Higgins, D. G.** (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* **23**(21), 2947-8.
- Lauring, A. S. y Andino, R.** (2010). Quasispecies theory and the behavior of RNA viruses. *PLoS Pathog* **6**(7), e1001005.
- Li, J., Browning, S., Mahal, S. P., Oelschlegel, A. M. y Weissmann, C.** (2010). Darwinian evolution of prions in cell culture. *Science* **327**(5967), 869-72.
- Li, Y. y Wilke, C. O.** (2004). Digital evolution in time-dependent fitness landscapes. *Artif Life* **10**(2), 123-34.
- Lopez-Galindez, C., Ortin, J., Domingo, E., del Rio, L., Perez-Brena, P. y Najera, R.** (1985). Heterogeneity among influenza H3N2 isolates recovered during an outbreak. Brief report. *Arch Virol* **85**(1-2), 139-44.
- Lorenzo-Redondo, R., Borderia, A. V. y Lopez-Galindez, C.** (2011). Dynamics of HIV-1 “in vitro” fitness recovery. *J Virol*.
- Lu, Y. L., Spearman, P. y Ratner, L.** (1993). Human immunodeficiency virus type 1 viral protein R localization in infected cells and virions. *J Virol* **67**(11), 6542-50.
- Lynch, M., Borger, R., Butcher, D. y Gabriel, W.** (1993). The mutational meltdown in asexual populations. *Journal of heredity* **84**, 339-344.
- Macken, C. A. y Perelson, A. S.** (1989). Protein evolution on rugged landscapes. *Proc Natl Acad Sci U S A* **86**(16), 6191-5.
- Malim, M. H.** (2009). APOBEC proteins and intrinsic resistance to HIV-1 infection. *Philos Trans R Soc Lond B Biol Sci* **364**(1517), 675-87.
- Mansky, L. M. y Temin, H. M.** (1995). Lower In Vivo Mutation Rate of Human Immunodeficiency Virus Type 1 than That Predicted from the Fidelity of Purified Reverse Transcriptase. *Journal of Virology* **69**, 5087-5094.
- Martell, M., Esteban, J. I., Quer, J., Genesca, J., Weiner, A., Esteban, R., Guardia, J. y Gomez, J.** (1992). Hepatitis C virus (HCV) circulates as a population of different but closely related genomes: quasispecies nature of HCV genome distribution. *J Virol* **66**(5), 3225-9.
- Masur, H., Michelis, M. A., Greene, J. B., Onorato, I., Stouwe, R. A., Holzman, R. S., Wormser, G., Brettman, L., Lange, M., Murray, H. W. y Cunningham-Rundles, S.** (1981). An outbreak of community-acquired *Pneumocystis carinii* pneumonia: initial manifestation of cellular immune dysfunction. *N Engl J Med* **305**(24), 1431-8.
- Maynard-Smith, J.** (1970). Natural Selection and the Concept of a Protein Space. *Nature* **225**, 563-564.
- McAllister, J. J., Phillips, D., Millhouse, S., Conner, J., Hogan, T., Ross, H. L. y Wigdahl, B.** (2000). Analysis of the HIV-1 LTR NF-kappaB-proximal Sp site III: evidence for cell type-specific gene regulation and viral replication. *Virology* **274**(2), 262-77.
- Meyer, B. E. y Malim, M. H.** (1994). The HIV-1 Rev trans-activator shuttles between the nucleus and the cytoplasm. *Genes Dev* **8**(13), 1538-47.
- Miller, M. D., Warmerdam, M. T., Gaston, I., Greene, W. C. y Feinberg, M. B.** (1994). The human immunodeficiency virus-1 nef gene product: a positive factor for viral infection and replication in primary lymphocytes and macrophages. *J Exp Med* **179**, 101-113.
- Mukai, T., Chigusa, S. I., Mettler, L. E. y Crow, J. F.** (1972). Mutation rate and dominance of genes affecting viability in *Drosophila melanogaster*. *Genetics* **72**(2), 335-55.
- Muller, H. J.** (1964). The relation of recombination to mutational advance. *Mutat Res* **1**, 2-9.
- Muñoz, A., Sabin, C. A. y Phillips, A. N.** (1997). The incubation period of AIDS. *AIDS* **11** (suppl. A), 69-76.
- Myers, G.** (1994). Tenth Anniversary Perspectives on AIDS. HIV: Between Past and Future. *AIDS Research and Human Retroviruses* **10**, 1317-1324.

- Nei, M. y Gojobori, T.** (1986). Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol* **3**(5), 418-26.
- Neil, S. J., Zang, T. y Bieniasz, P. D.** (2008). Tetherin inhibits retrovirus release and is antagonized by HIV-1 Vpu. *Nature* **451**(7177), 425-30.
- Nerrienet, E., Santiago, M. L., Foupouapouognigni, Y., Bailes, E., Mundy, N. I., Njinku, B., Kfutwah, A., Muller-Trutwin, M. C., Barre-Sinoussi, F., Shaw, G. M., Sharp, P. M., Hahn, B. H. y Ayouba, A.** (2005). Simian Immunodeficiency Virus Infection in Wild-Caught Chimpanzees from Cameroon. *J. Virol.* **79**(2), 1312-1319.
- Novella, I. S.** (2003). Contributions of vesicular stomatitis virus to the understanding of RNA virus evolution. *Current Opinion in Microbiology* **6**(4), 399-405.
- Novella, I. S., Duarte, E. A., Elena, S. F., Moya, A., Domingo, E. y Holland, J. J.** (1995). Exponential increases of RNA virus fitness during large population transmissions. *Proceedings of the National Academy of Sciences of the United States of America* **92**, 5841-5844.
- Nowak, M. A.** (2006). "Evolutionary dynamics: exploring the equations of life." Harvard University Press
- Ojosnegros, S., Beerenwinkel, N., Antal, T., Nowak, M. A., Escarmis, C. y Domingo, E.** (2010). Competition-colonization dynamics in an RNA virus. *Proc Natl Acad Sci U S A* **107**(5), 2108-12.
- Olivares, I., Shaw, G. y Lopez-Galindez, C.** (1997). Phenotypic switch in a spanish HIV Type 1 isolate on serial passage on MT-4 cells. *AIDS Res Hum Retroviruses* **13** (11), 979-84.
- Pawlotsky, J. M.** (2006). Hepatitis C virus population dynamics during infection. *Curr Top Microbiol Immunol* **299**, 261-84.
- Perales, C., Lorenzo-Redondo, R., López-Galíndez, C., Martínez, M. A. y Domingo, E.** (2010). Mutant spectra in virus behavior. *Future Virology* **5**, 679-698.
- Perales, C., Mateo, R., Mateu, M. G. y Domingo, E.** (2007). Insights into RNA virus mutant spectrum and lethal mutagenesis events: replicative interference and complementation by multiple point mutants. *J Mol Biol* **369**(4), 985-1000.
- Perelson, A. S., Neumann, A. U., Markowitz, M., Leonard, J. M. y Ho, D. D.** (1996). HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time. *Science* **271**(5255), 1582-6.
- Perucho, M., Goldfarb, M., Shimizu, K., Lama, C., Fogh, J. y Wigler, M.** (1981). Human-tumor-derived cell lines contain common and different transforming genes. *Cell* **27**(3 Pt 2), 467-76.
- Pitt, J. N. y Ferre-D'Amare, A. R.** (2010). Rapid construction of empirical RNA fitness landscapes. *Science* **330**(6002), 376-9.
- Plantier, J. C., Leoz, M., Dickerson, J. E., De Oliveira, F., Cordonnier, F., Lemee, V., Damond, F., Robertson, D. L. y Simon, F.** (2009). A new human immunodeficiency virus derived from gorillas. *Nat Med* **15**(8), 871-2.
- Pommier, Y., Johnson, A. A. y Marchand, C.** (2005). Integrase inhibitors to treat HIV/Aids. *Nature Reviews Drug Discovery* **4**, 236-248.
- Pond, S. L., Frost, S. D. y Muse, S. V.** (2005). HyPhy: hypothesis testing using phylogenies. *Bioinformatics* **21**(5), 676-9.
- Poon, B., Grovit-Ferbas, K., Stewart, S. A. y Chen, I. S. Y.** (1998). Cell cycle arrest by Vpr in HIV-1 virions and insensitivity to antiretroviral agents. *Science* **281**(5374), 266-9.
- Posada, D.** (2008). jModelTest: phylogenetic model averaging. *Mol Biol Evol* **25**(7), 1253-6.
- Quinones-Mateu, M. E. y Arts, E. J.** (2006). Virus fitness: concept, quantification, and application to HIV population dynamics. *Curr Top Microbiol Immunol* **299**, 83-140.
- Ratner, L., Gallo, R. C. y Wong-Staal, F.** (1985). HTLV-III, LAV, ARV are variants of same AIDS virus. *Nature* **313**(6004), 636-7.
- Rouzine, I. M., Rodrigo, A. y Coffin, J. M.** (2001). Transition between Stochastic Evolution and Deterministic Evolution in the Presence of Selection: General Theory and Application to Virology. *Microbiol. Mol. Biol. Rev.* **65**(1), 151-185.

- Ruiz-Jarabo, C. M., Arias, A., Baranowski, E., Escarmis, C. y Domingo, E. (2000). Memory in Viral Quasispecies, Vol. 74, pp. 3543-3547.
- Salazar-Gonzalez, J. F., Bailes, E., Pham, K. T., Salazar, M. G., Guffey, M. B., Keele, B. F., Derdeyn, C. A., Farmer, P., Hunter, E., Allen, S., Manigart, O., Mulenga, J., Anderson, J. A., Swanstrom, R., Haynes, B. F., Athreya, G. S., Korber, B. T. M., Sharp, P. M., Shaw, G. M. y Hahn, B. H. (2008). Deciphering Human Immunodeficiency Virus Type 1 Transmission and Early Envelope Diversification by Single-Genome Amplification and Sequencing. *En "J.Virol"*, Vol. 82, pp. 3952-3970.
- Salazar-Gonzalez, J. F., Salazar, M. G., Keele, B. F., Learn, G. H., Giorgi, E. E., Li, H., Decker, J. M., Wang, S., Baalwa, J., Kraus, M. H., Parrish, N. F., Shaw, K. S., Guffey, M. B., Bar, K. J., Davis, K. L., Ochsenbauer-Jambor, C., Kappes, J. C., Saag, M. S., Cohen, M. S., Mulenga, J., Derdeyn, C. A., Allen, S., Hunter, E., Markowitz, M., Hraber, P., Perelson, A. S., Bhattacharya, T., Haynes, B. F., Korber, B. T., Hahn, B. H. y Shaw, G. M. (2009). Genetic identity, biological phenotype, and evolutionary pathways of transmitted/founder viruses in acute and early HIV-1 infection. *En "J.Exp.Med."* pp. jem.20090378.
- Salemi, M., Strimmer, K., Hall, W. W., Duffy, M., Delaporte, E., Mboup, S., Peeters, M. y Vandamme, A. M. (2001). Dating the common ancestor of SIVcpz and HIV-1 group M and the origin of HIV-1 subtypes using a new method to uncover clock-like molecular evolution. *Faseb J* 15(2), 276-8.
- Sanchez-Palomino, S., Rojas, J. M., Martinez, M. A., Fenyő, E. M., Najera, R., Domingo, E. y Lopez-Galindez, C. (1993). Dilute passage promotes expression of genetic and phenotypic variants of human immunodeficiency virus type 1 in cell culture. *J Virol* 67 (5), 2938-43.
- Sattentau, Q. J., Moore, J. P., Vignaux, F., Traincard, F. y Poignard, P. (1993). Conformational changes induced in the envelope glycoproteins of the human and simian immunodeficiency viruses by soluble receptor binding. *J Virol* 67(12), 7383-93.
- Sauder, C. J., Vandenburgh, K. M., Iskow, R. C., Malik, T., Carbone, K. M. y Rubin, S. A. (2006). Changes in mumps virus neurovirulence phenotype associated with quasispecies heterogeneity. *Virology* 350(1), 48-57.
- Sayle, R. A. y Milner-White, E. J. (1995). RASMOL: biomolecular graphics for all. *Trends Biochem Sci* 20(9), 374.
- Schrofelbauer, B., Yu, Q. y Landau, N. R. (2004). New insights into the role of Vif in HIV-1 replication. *AIDS Rev* 6(1), 34-9.
- Schwartz, O., Marechal, V., Le Gall, S., Lemonnier, F. y Heard, J. M. (1996). Endocytosis of major histocompatibility complex class I molecules is induced by the HIV-1 Nef protein. *Nat Med* 2(3), 338-42.
- Schwartz, O., Maržchal, V., Danos, O. y Heard, a. J. M. (1995). Human immunodeficiency virus type1 nef increases the efficiency of reverse transcription in the infected cell. *J. Virol.* 69 (7), 4053-9.
- Shankarappa, R., Margolick, J. B., Gange, S. J., Rodrigo, A. G., Upchurch, D., Farzadegan, H., Gupta, P., Rinaldo, C. R., Learn, G. H., He, X., Huang, X. L. y Mullins, J. I. (1999). Consistent viral evolutionary changes associated with the progression of human immunodeficiency virus type 1 infection. *J Virol* 73(12), 10489-502.
- Simon, F., Mauciere, P., Roques, P., Loussert-Ajaka, I., Muller-Trutwin, M. C., Saragosti, S., Georges-Courbot, M. C., Barre-Sinoussi, F. y Brun-Vezinet, F. (1998). Identification of a new human immunodeficiency virus type 1 distinct from group M and group O. *Nat Med* 4(9), 1032-7.
- Sobrino, F., Davila, M., Ortin, J. y Domingo, E. (1983). Multiple Genetic Variants Arise in the Course of Replication of Foot-and-Mouth Disease Virus in Cell Culture. *Virology* 128, 310-318.

- Spina, C. A., Kwoh, T. J., Chowes, M. Y., Guatelli, J. C. y Richman, D. D.** (1994). The importance of nef in the induction of human immunodeficiency virus type 1 replication from primary quiescent CD4 lymphocytes. *J Exp Med* **179**, 115-23.
- Strebel, K., Daugherty, D., Clouse, K., Cohen, D., Folks, T. y Martin, M. A.** (1987). The HIV 'A' (sor) gene product is essential for virus infectivity. *Nature* **328**(6132), 728-30.
- PAUP*:** phylogenetic analysis using parsimony (* and other methods) Version 4.0 bets. Sinauer, Sunderland, Mass.
- Tachedjian, G., French, M. y Mills, J.** (1998). Coresistance to zidovudine and foscarnet is associated with multiple mutations in the human immunodeficiency virus type 1 reverse transcriptase. *Antimicrob Agents Chemother* **42**(11), 3038-43.
- Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. y Kumar, S.** MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Mol Biol Evol*.
- Tejero, H., Marin, A. y Montero, F.** (2011). The relationship between the error catastrophe, survival of the flattest, and natural selection. *BMC Evol Biol* **11**, 2.
- Temin, H. M. y Mizutani, S.** (1970). RNA-dependent DNA polymerase in virions of Rous sarcoma virus. *Nature* **226**(5252), 1211-3.
- Thali, M., Moore, J. P., Furman, C., Charles, M., Ho, D. D., Robinson, J. y Sodroski, J.** (1993). Characterization of conserved human immunodeficiency virus type 1 gp120 neutralization epitopes exposed upon gp120-CD4 binding. *J Virol* **67**(7), 3978-88.
- Troyer, R. M., Collins, K. R., Abraha, A., Fraundorf, E., Moore, D. M., Krizan, R. W., Toossi, Z., Colebunders, R. L., Jensen, M. A., Mullins, J. I., Vanham, G. y Arts, E. J.** (2005). Changes in Human Immunodeficiency Virus Type 1 Fitness and Genetic Diversity during Disease Progression. *J. Virol.* **79**(14), 9006-9018.
- Vandamme, A.-M.** (2009). Basic concepts of molecular evolution. En "The Phylogenetic Handbook. A Practical Approach to Phylogenetic Analysis and Hypothesis Testing. 2nd Edition" (P. Lemey, M. Salemi, and A.-M. Vandamme, Eds.), pp. 3-29. Cambridge University Press.
- Vignuzzi, M., Stone, J. K., Arnold, J. J., Cameron, C. E. y Andino, R.** (2006). Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population. *Nature* **439**(7074), 344-8.
- Vignuzzi, M., Wendt, E. y Andino, R.** (2008). Engineering attenuated virus vaccines by controlling replication fidelity. *Nat Med* **14**(2), 154-61.
- von Schwedler, U., Kornbluth, R. S. y Trono, D.** (1994). The nuclear localization signal of the matrix protein of human immunodeficiency virus type 1 allows the establishment of infection in macrophages and quiescent T lymphocytes. *Proc Natl Acad Sci U S A* **91**(15), 6992-6.
- Wain-Hobson, S.** (1992). Human immunodeficiency virus type 1 quasispecies in vivo and ex vivo. *Curr Top Microbiol Immunol* **176**, 181-93.
- Wain-Hobson, S., Sonigo, P., Danos, O., Cole, S. y Alizon, M.** (1985). Nucleotide sequence of the AIDS virus, LAV. *Cell* **40**(1), 9-17.
- Walker, B. D. y Burton, D. R.** (2008). Toward an AIDS vaccine. *Science* **320**(5877), 760-4.
- White, J. M.** (1992). Membrane Fusion. *Science* **258**, 917-924.
- Wilke, C. O.** (2005). Quasispecies theory in the context of population genetics. *BMC Evol Biol* **5**, 44.
- Wolinsky, S. M., Korber, B. T., Neumann, A. U., Daniels, M., Kunstman, K. J., Whetsell, A. J., Furtado, M. R., Cao, Y., Ho, D. D. y Safrin, J. T.** (1996). Adaptive evolution of human immunodeficiency virus-type 1 during the natural course of infection. *Science* **272**(5261), 537-42.
- Wong, W. S., Yang, Z., Goldman, N. y Nielsen, R.** (2004). Accuracy and power of statistical methods for detecting adaptive evolution in protein coding sequences and for identifying positively selected sites. *Genetics* **168**(2), 1041-51.

- Worobey, M., Gemmel, M., Teuwen, D. E., Haselkorn, T., Kunstman, K., Bunce, M., Muyembe, J. J., Kabongo, J. M., Kalengayi, R. M., Van Marck, E., Gilbert, M. T. y Wolinsky, S. M.** (2008). Direct evidence of extensive diversity of HIV-1 in Kinshasa by 1960. *Nature* **455**(7213), 661-4.
- Wright, S.** (1932). The roles of mutation, inbreeding, crossbreeding, and selection in evolution. *Proceedings of the Sixth International Congress on Genetics*, 355-366.
- Yuan, X., Yu, X., Lee, T. H. y Essex, M.** (1993). Mutations in the N-terminal region of human immunodeficiency virus type 1 matrix protein block intracellular transport of the Gag precursor. *J Virol* **67**(11), 6387-94.
- Yuste, E., Borderia, A. V., Domingo, E. y Lopez-Galindez, C.** (2005). Few Mutations in the 5' Leader Region Mediate Fitness Recovery of Debilitated Human Immunodeficiency Type 1 Viruses. *J. Virol.* **79**(9), 5421-5427.
- Yuste, E., Lopez-Galindez, C. y Domingo, E.** (2000). Unusual distribution of mutations associated with serial bottleneck passages of human immunodeficiency virus type 1. *J Virol* **74**(20), 9546-52.
- Yuste, E., Sanchez-Palomino, S., Casado, C., Domingo, E. y Lopez-Galindez, C.** (1999). Drastic fitness loss in human immunodeficiency virus type 1 upon serial bottleneck events. *J Virol* **73** (4), 2745-51.
- Zhou, W., Parent, L. J., Wills, J. W. y Resh, M. D.** (1994). Identification of a membrane-binding domain within the amino-terminal region of human immunodeficiency virus type 1 Gag protein which interacts with acidic phospholipids. *J Virol* **68**(4), 2556-69.
- Zhu, H., Jian, H. y Zhao, L. J.** (2004). Identification of the 15FRFG domain in HIV-1 Gag p6 essential for Vpr packaging into the virion. *Retrovirology* **1**, 26.
- Zhu, T., Korber, B. T., Nahmias, A. J., Hooper, E., Sharp, P. M. y Ho, D. D.** (1998). An African HIV-1 sequence from 1959 and implications for the origin of the epidemic. *Nature* **391**, 594-597.

Trabajos Publicados

Initial Fitness Recovery of HIV-1 Is Associated with Quasispecies Heterogeneity and Can Occur without Modifications in the Consensus Sequence

Antonio V. Bordería^{1,2}, Ramon Lorenzo-Redondo^{1,2}, Maria Pernas¹, Concepción Casado¹, Tamara Alvaro¹, Esteban Domingo^{2,3}, Cecilio Lopez-Galindez^{1*}

1 Virología Molecular, Centro Nacional de Microbiología (CNM), Instituto de Salud Carlos III, Majadahonda, Madrid, Spain, **2** Centro de Biología Molecular “Severo Ochoa” (CSIC-UAM), Madrid, Spain, **3** Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBERehd), Barcelona, Spain

Abstract

Background: Fitness recovery of HIV-1 “in vitro” was studied using viral clones that had their fitness decreased as a result of plaque-to-plaque passages.

Principal Findings: After ten large population passages, the viral populations showed an average increase of fitness, although with wide variations among clones. While 5 clones showed significant fitness increases, 3 clones showed increases that were only marginally significant ($p < 0.1$), and 4 clones did not show any change. Fitness recovery was not accompanied by an increase in p24 production, but was associated with an increase in viral titer. Few mutations (an average of 2 mutations per genome) were detected in the consensus nucleotide sequence of the entire genome in all viral populations. Five of the populations did not fix any mutation, and three of them displayed marginally significant fitness increases, illustrating that fitness recovery can occur without detectable alterations of the consensus genomic sequence. The investigation of other possible viral factors associated with the initial steps of fitness recovery, showed that viral quasispecies heterogeneity increased between the initial clones and the passaged populations. A direct statistical correlation between viral heterogeneity and viral fitness was obtained.

Conclusions: Thus, the initial fitness recovery of debilitated HIV-1 clones was mediated by an increase in quasispecies heterogeneity. This observation, together with the invariance of the consensus sequence despite fitness increases demonstrates the relevance of quasispecies heterogeneity in the evolution of HIV-1 in cell culture.

Citation: Bordería AV, Lorenzo-Redondo R, Pernas M, Casado C, Alvaro T, et al. (2010) Initial Fitness Recovery of HIV-1 Is Associated with Quasispecies Heterogeneity and Can Occur without Modifications in the Consensus Sequence. PLoS ONE 5(4): e10319. doi:10.1371/journal.pone.0010319

Editor: Srikanth Tripathy, National AIDS Research Institute, India

Received: November 19, 2009; **Accepted:** March 29, 2010; **Published:** April 26, 2010

Copyright: © 2010 Bordería et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: Work in the Centro Nacional de Microbiología was supported by grants, SAF 2005/03833, SAF 2007/61036 by the Plan Nacional del SIDA and in part by the Red Temática Cooperativa de investigación en SIDA (Red de grupos 173) of the Fondo de Investigaciones Sanitarias de la Seguridad Social (FISs) and Fundación para la Investigación y la Prevención del SIDA en España (FIPSE) grant 36558/06, 36641/07, 36779/08, 36641/07 and 360766/09. Work in the Centro de Biología Molecular “Severo Ochoa” was supported by grant BFU 2008-02816/BMC from Ministerio de Ciencia e Innovación, FIPSE 36558/06 and 360766/09 and Fundación Ramón Areces. Centro de Investigaciones Biomédicas en Red (CIBER): enfermedades hepáticas y digestivas (ehd) is funded by Instituto de Salud Carlos III. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: clopez@isciii.es

These authors contributed equally to this work.

Introduction

Experiments of virus evolution in cell culture have been employed to study the relevance of genetic concepts for RNA viruses (including Muller’s ratchet, the Red Queen hypothesis, genetic drift in bottleneck transfers, among others; as review see [1]). Such studies have been also extremely useful in the understanding of the generation and quantification of genetic variation of human immunodeficiency virus type 1 (HIV-1) [2,3,4]. These investigations showed that HIV-1 populations are composed of swarms of related variants, often centered around a high frequency variant or master sequence, forming what is known as viral quasispecies [5,6]. According to the theoretical quasispecies model, proposed by Eigen [7], viral evolution does not operate at the level of the individual genome but rather the target of selection is the quasispecies as a whole [6].

Recently, this concept has been experimentally demonstrated for RNA viruses with the implication of quasispecies diversity in the pathogenesis of poliovirus [8,9], and other RNA viruses [10]. In addition, the pathogenicity of a Mumps vaccine strain has been related to the quasispecies heterogeneity [11]. High fidelity viral strains that produce less heterogeneous and less pathogenic viral populations are under study as new poliovirus vaccine candidates [12].

The viral population size changes during natural HIV-1 infections. The transmission to a new individual often implies a reduction in the population size, known as transmission bottleneck, which affects the composition of the quasispecies in the recipient host [13]. Once the infection is successful in the recipient host, viruses replicate generating diversity and expanding their populations to explore additional regions of the sequence space.

In a similar way, antiretroviral treatments, such as highly active antiretroviral therapy (HAART), dramatically reduce the viral load producing a population bottleneck that may be shaping virus evolution. In previous studies “*in vitro*” by our group, we documented a rapid HIV-1 fitness decrease after serial bottlenecks passages [3], and also rapid fitness recovery attained when the virus was subjected to large population passages [3,4,14,15,16]. Furthermore, we found that a very limited number of mutations in the consensus HIV-1 genomic sequences was responsible for the fitness changes [4].

To understand the variables that affect fitness recovery, ten HIV-1 clones (two of which were passaged in duplicate), whose capacity to produce infectious progeny had decreased by plaque-to-plaque transfers, were subjected to ten large population passages. The molecular changes associated with fitness recovery were analyzed by comparing the consensus genomic nucleotide sequence as well as the sequences of individual molecular clones. We detected fixation of mutations in the consensus sequences only in 7 out of the 12 passaged viruses. Heterogeneity analysis of the

mutant spectra in the viral populations showed an increase of complexity along the passages and an overall positive correlation between quasispecies heterogeneity and fitness increase. These results provide the first evidence that in the course of HIV-1 evolution, mutant spectrum complexity can be a determinant of virus behavior.

Results

HIV-1 fitness increase after ten large population passages

Fitness of HIV-1 viral clones previously subjected to serial plaque-to-plaque transfers [3] was recovered after ten large population passages (Figure 1). Viral fitness was determined using J1 clone as reference [4]. The initial clones showed a mean viral fitness of 0.5418 (standard error (SEM) ± 0.054), whereas for the corresponding passaged populations, the average fitness was 0.795 ± 0.037 , a difference that was statistically significant (Wilcoxon signed rank test; $p = 0.0034$, Figure 2). This result indicates that few large population passages are sufficient to

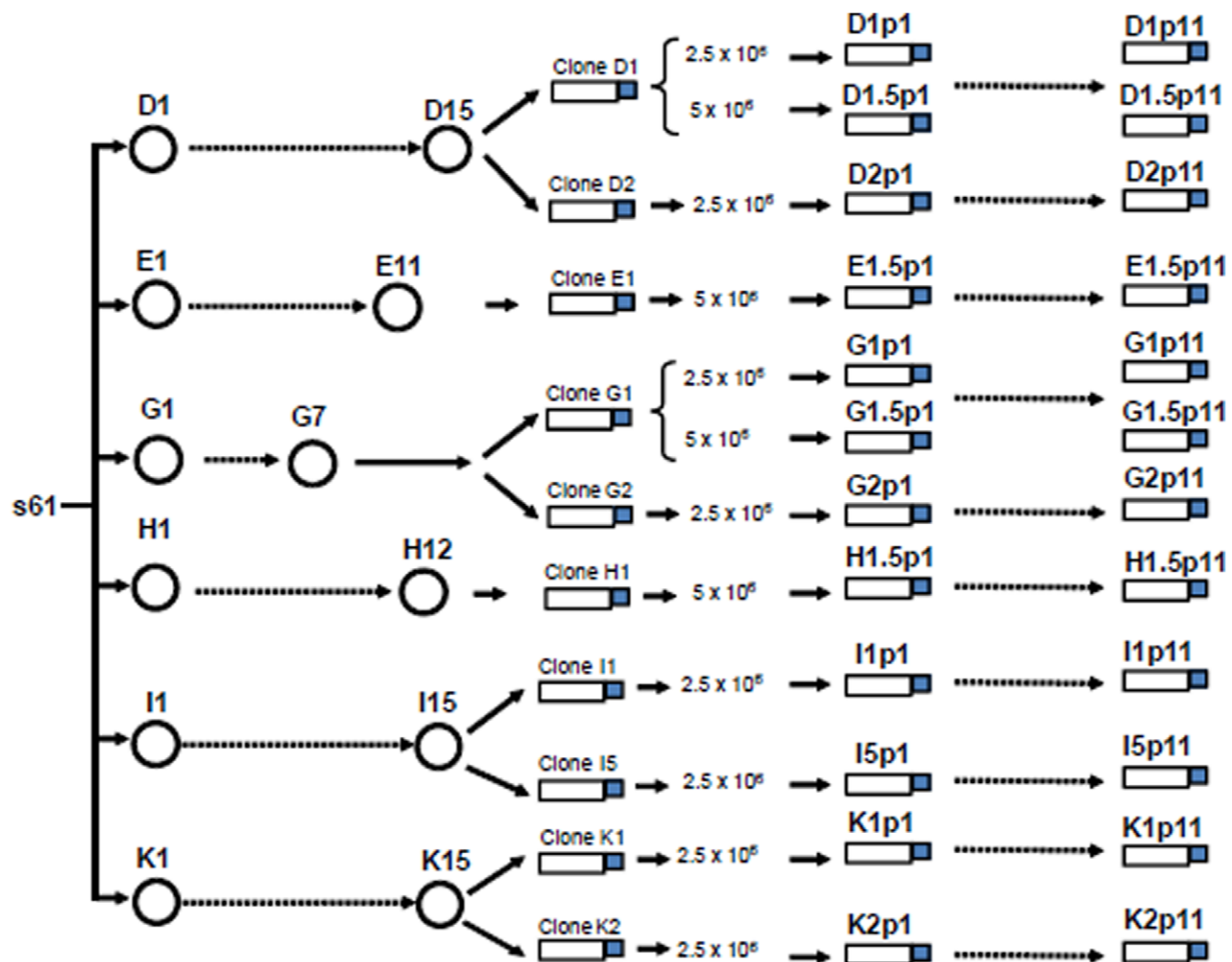


Figure 1. Genealogy of the viral clones studied. Representation of the serial plaque-to-plaque transfers (circles) and large population passages (bottles and dotted arrows) in 2.5×10^6 and 5×10^6 MT-4 cells performed with HIV-1 clones. The experimental procedures and the origins of natural isolate s61 and of the different clones are described in references [2,3] and in Materials and methods. HIV-1 clones were obtained from the populations after the plaque-to-plaque transfers, and are named according to the number of large population recovery passages. Viral populations are indicated by letters followed by a number that identifies the clone used, followed by p1 for the initial and p11 for the final passage. For clones D1, G1, E1 and H1, number 5 after the dot indicates that these clones were passaged by infecting 5×10^6 MT-4 cells. doi:10.1371/journal.pone.0010319.g001

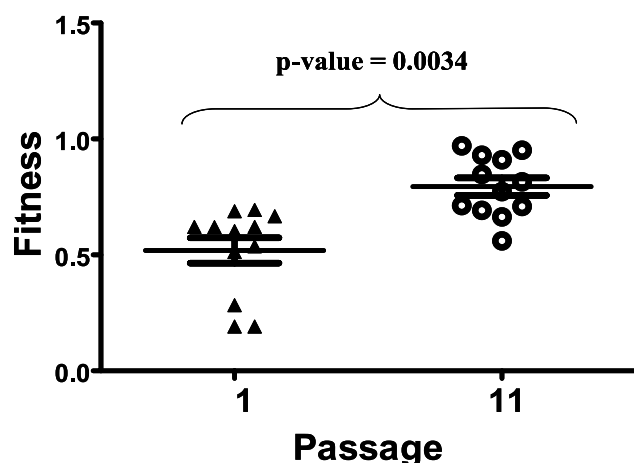


Figure 2. Fitness increase after ten serial passages of HIV-1 clones. Fitness values were obtained from values of each clone at passage 1 (▲) and 11 (○) and represented together with the mean value and standard errors. A Wilcoxon signed rank test was used to compare the values; the p-value of this test is shown above the points (Sum of signed ranks = -70). doi:10.1371/journal.pone.0010319.g002

increase viral fitness of HIV clones that had undergone bottleneck passages. The increase in fitness, however, varied substantially among clones, with increases ranging from 386% to undetectable (0%) (Table 1). Five of the clones showed increases that were statistically significant at $p < 0.05$; 3 clones were marginally significant (at $p < 0.10$), and in 4 clones there was no fitness gain.

Viral infectivity and p24 production in HIV-1 clones

To identify viral factors that could be responsible for the fitness increase, viral titers and p24 protein levels were quantified for all HIV-1 clones before and after the large population passages. Passages resulted in a statistically significant increase in viral titer (Wilcoxon signed rank test; p -value = 0.021) (Figure 3a). To test whether this result was due to an increase in the virion production, viral p24 protein was quantified in the cell culture supernatants. No differences in the protein levels were observed during the passages (Wilcoxon signed rank test; p -value = 0.2334) (Figure 3b). These results suggest that the observed increase in viral titer is due to an increase of the virus infectivity, and not to an increase in virion production.

Detection of mutations in the consensus nucleotide sequences

Because mutations are generally considered the principal cause of changes in viral fitness and other phenotypic alterations [17,18,19,20,21,22,23], we compared the consensus nucleotide sequence of each clone before and after being subjected to large population passages. Few mutations were fixed in each viral clone following large population passages (average of 2 mutations per genome, range 0 to 10) (Table 1). Only 7 out of the 12 populations studied showed changes in their genomic sequence, which included four reversions (17% of the changes), one mutation that eliminated a stop codon in *nef* coding region of virus H (Table 1) and a common mutation in the 5'LTR in three viruses, as described [4]. The results confirm, in agreement with our previous work, that few mutations in the consensus sequence can mediate fitness recovery [4].

Of the 7 clones showing accumulation of mutations, five displayed statistically significant fitness increases, whereas two

clones (G1.5p11 and G2p11) did not show any significant fitness increase (Table 1). In addition, there were clones that acquired several mutations (such as clone E1) and manifested a lower fitness increase than clones that acquired a single mutation (clones in lineage D). Strikingly, clones G1p11, I1p11 and K1p11, that did not include any mutation in the consensus sequence, displayed an increase in fitness (with $p = 0.0664$, $p = 0.0565$ and $p = 0.0963$ respectively in the Graphpad slope comparison test). Therefore, fitness increase was not directly correlated with the number of mutations in the consensus sequence, and in some clones, this fitness increase could not be linked to any mutation in the consensus sequence of the population.

Quasispecies heterogeneity and viral fitness

To investigate viral factors, other than dominant mutations that could be involved in fitness increase, we quantified the heterogeneity of mutant spectra, by sequencing molecular clones of four genomic regions of the initial clones and the corresponding passaged populations. At least 20 clones were analyzed for each of the four regions, which spans 2,005 nucleotides per clone, and thus represents 480,144 nucleotides. The distribution of mutations among the clones and genomic regions is described in Table 2. The analysis documented that there was a significant difference in heterogeneity (Wilcoxon signed rank test $p = 0.0122$), quantified as the sum of Hamming average distances in each region, between the initial (5.546 ± 0.1481 substitutions) and the passaged viral populations (6.464 ± 0.33 substitutions) (Figure 4).

In addition, a global correlation analysis indicated a significant correlation between viral heterogeneity and viral fitness with a p -value = 0.05 (Table 1 and Figure 5). Therefore, fitness increase of HIV-1 can occur without modifications in the consensus sequence of the replicating quasispecies and quasispecies complexity is associated with the increases of HIV-1 fitness.

Discussion

The present study was designed to investigate the early steps of fitness recovery of ten debilitated HIV-1 clones subjected to only ten serial large population passages. In the course of fitness gain, the genetic variation differed from clone to clone, but a global diversification of viral quasispecies was observed, illustrating that the initial phase of fitness recovery is characterized by the generation of variation. The increase of fitness varied among clones, with higher increases in clones with lower initial values (D1, D1.5 and D2). This fitness augmentation was not related to viral production, as measured by p24 levels, but it was associated with an increase in viral infectivity (Figure 3). Consensus nucleotide sequence of the entire viral genome showed the accumulation of mutations in some passaged populations but not in others. This result documents that fitness alterations can occur without variation of the consensus sequence. Similar results have been previously reported for several other riboviruses [8,9,24,25,26,27]. The global fitness correlation with heterogeneity could be differentiated between non-significant correlation in the initial populations and a statistical association in passage 11 viruses (see Figure 5). Fitness increase correlated with quasispecies heterogeneity but not with the number of mutations fixed in the viral populations (Table 1). It must be noted that the experiment was designed to test the initial process of fitness recovery, and that it cannot be excluded that additional passages might result in further fitness increases accompanied by modifications of the consensus sequence. The key point of this cell culture study is that a transient fitness increase can occur without being reflected in the consensus sequence and that critical information in HIV-1

Table 1. Viral clones, fitness increases and mutations fixed in the consensus sequences.

Viral Clones								
Initial	Final	Fitness (SEM)	Increase(%)	p-value	Number of mutations	Localization ^a	Change of nucleotide	Change of aa
D1p1		0,2 (0,1)						
	D1p11	0,9 (0,06)	376	0.0003	1	U5 5' LTR (593)	G à T	NA
	D1.5p11^d	0,9 (0,07)	386	0.0013	1	U5 5' LTR (593)	G à T	NA
D2p1		0,3 (0,03)						
	D2p11	1,0 (0,03)	243	<0,0001	1	U5 5' LTR (593)	G à T	NA
E1.5p1		0,7 (0,04)						
	E1.5p11^d	0,8 (0,03)	29	0.0017	10	U3 5'LTR (255)	G à A	NA
						U5 5'LTR (572)	T à C	NA
						env (2293)	G à A ^b	D249N (gp41)
						rev (197)		R66Q (rev)
						rev (209)	T à C	L70P
						rev (349)	T à C	STOP117Q
						env (2580)	T à A	Synonymous
						nef (426)	C à T	Synonymous
						nef (436)	C à G	V146L
						nef (544)	G à A	V182I
						nef (550)	A à G	R184G
G1p1		0,6 (0,05)						
	G1p11	0,8 (0,03)	25	0.0664	0			
	G1.5p11^d	0,7 (0,02)	15	0.2329	4	pol (1054)	A à G	K194E (RT)
						pol (2346)	T à A	D64E (IN)
						pol (2767)	G à A	A205T (IN)
						rev (100)	A à G ^b	T34A (rev)
						tat (239)		D80G (tat)
G2p1		0,7 (0,03)						
	G2p11	0,7 (0,01)	3	0.7815	4	pol (1060)	G à A	E196K ^c (RT)
						pol (1062)	G à A	
						pol (1063)	C à A	Q197K ^c (RT)
						pol (1065)	G à A	
H1.5p1		0,5 (0,04)						
	H1.5p11^d	1,0 (0,01)	85	<0,0001	3	gag (654)	G à A	M86I (p24)
						env (1549)	A à G	T1A (gp41)
						nef (117)	-à A	Phase recovery
I1p1		0,5 (0,08)						
	I1p11	0,7 (0,02)	29	0.0565	0	-	-	-
I5p1		0,6 (0,03)						
	I5p11	0,6 (0,03)	-7	0.6645	0	-	-	-
K1p1		0,7 (0,004)						
	K1p11	0,8 (0,03)	17	0.0963	0	-	-	-
K2p1		0,7 (0,04)						
	K2p11	0,7 (0,03)	0	0.9494	0	-	-	-
Total					24			

^aLocation of mutations is assigned taking HIV-1 strain HXB-2 as a reference (Accession number #K03455). Mutations are numbered according to their position in the gene. NA: not applicable. Reversions are underlined. Amino acid substitutions are numbered in reference to the protein affected.

^bNucleotide substitution producing an amino acid change in two proteins.

^cBoth substitutions lead to the same amino acid change.

^dViruses passaged in 5·10⁶ MT-4 cells are doubly underlined.

doi:10.1371/journal.pone.0010319.t001

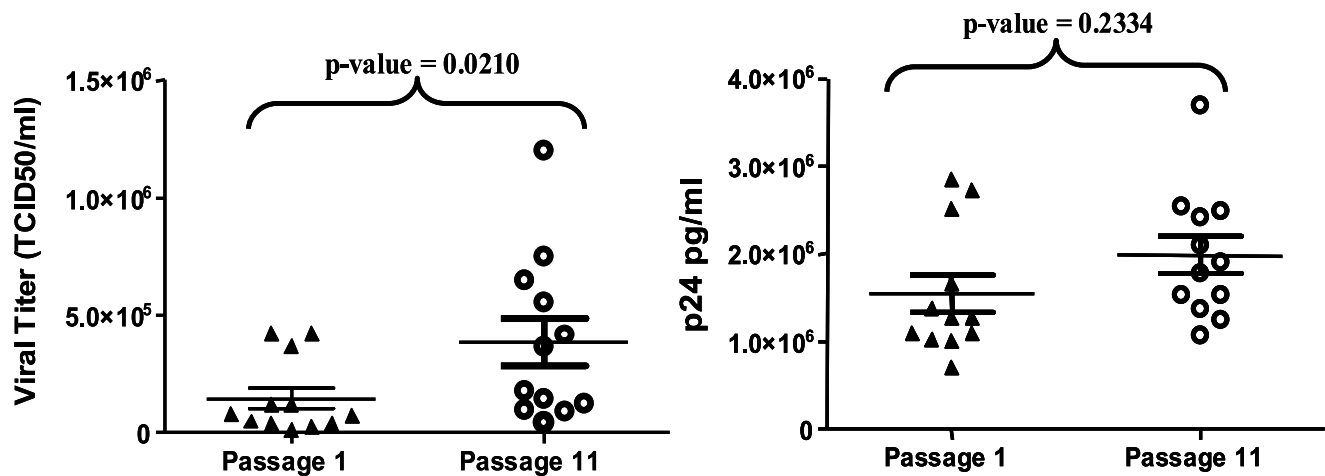


Figure 3. Virological characterization of HIV-1 clones in the initial and final passages. A) Viral titers of populations at passage 1 and 11 were obtained by infecting MT-2 cell with the corresponding supernatants, and are expressed as TCID₅₀ (see Materials and methods). To compare the mean values, a Wilcoxon signed rank test was used (sum of signed ranks = -58) and the p-value is given above the data points. B) p24 protein was quantified using an indirect ELISA, as described in Materials and methods. Mean values of passages 1 and 11 are shown as pg/ml, and were used to compare both passages. The Wilcoxon p-value is displayed above the points (sum of signed ranks = -32). Symbols are as in Figure 2. doi:10.1371/journal.pone.0010319.g003

behavior can be lost if no analyses of mutant spectra are performed.

In the case of HIV-1, fitness change without variation in the consensus genomic sequence can be significant in the clinical set-up, because changes in highly relevant phenotypes, for example those associated with antiviral resistance, are generally monitored by the appearance of mutations in the consensus sequence of the viral populations [28]. The role of quasispecies heterogeneity in fitness increase was most evident in the populations that did not display changes in the consensus sequence. However; it may be also operating, at a hidden level of the mutant spectra, as suggested by the comparison of heterogeneity values, in populations that displayed a fitness increase and that incorporated mutations in their consensus sequences.

The experiment reported here was undertaken to simulate population changes likely to occur during HIV-1 natural

infections. High HIV-1 population sizes are generally attained at primo-infection, and virus replication is partially controlled by the immune system, primarily by the cytotoxic T cell (CTL) response [29,30,31,32,33,34,35], and later by neutralizing antibodies [36]. Consequently, the viral load drops to a viral set point, which remains stable during the asymptomatic phase, and is variable among patients. Genetic variability studies in HIV-1 patients [37,38,39] and in a set of typical progressors have established consistent patterns of viral heterogeneity and divergence from the initial transmission to AIDS [40].

The present study with HIV-1 reinforces the adequacy of quasispecies theory to interpret retrovirus evolution. The key observation is that fitness recovery cannot only be mediated by the fixation and dominance of advantageous mutations but also by an increase in quasispecies heterogeneity. Thus, the viral population as a whole is the one involved in the fitness recovery. This result supports the concept that the target of selection is not a single genome but the complete quasispecies [5,41,42]. Dominance of specific mutants is not a requirement for fitness gain, and internal interactions within quasispecies can affect retroviral evolution.

Materials and Methods

Cells, viruses and biological cloning

The HIV-1 parental population was isolate s61 [2], obtained by standard co-culture procedures, from a 4-year-old child with AIDS symptoms [2]. From this viral population, six biological clones D, G, E, H, I, and K were derived by plating on MT-4 cells [43]. These clones were subjected to serial plaque-to-plaque transfers to obtain populations D15, E11, G7, H12, I15 and K15 [3] (Figure 1). From these viral populations, two clones were obtained from lineages D, G, K, and I, and only one clone could be recovered from lineages E and H (Figure 1).

Large population passages of viruses

Clones D1, D2, E1, G1, G2, H1, I1, I5, K1, and K2 were subjected to ten large population passages by infecting 5×10^6 or 2.5×10^6 MT-4 cells at a multiplicity of infection (M.O.I.) of about 0.01 TCID₅₀/cell (Figure 1). Viral titers were obtained in MT-2

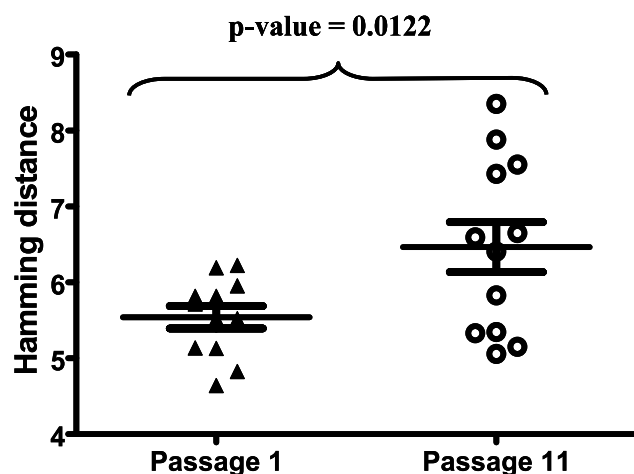


Figure 4. Global heterogeneity of the HIV-1 viral quasispecies at passages 1 and 11. Heterogeneity was measured as Hamming distance. Wilcoxon p-value to compare means is shown above the dots (sum of the signed ranks = -62). Symbols are as in Figure 2. doi:10.1371/journal.pone.0010319.g004

Table 2. Quasispecies heterogeneity per region and clone.

	LTR-gag		vpu		env(V1–V2)		env(V3–V4)	
PASSAGE	1	11	1	11	1	11	1	11
CLONE	Mean ± SEM	Mean ± SEM	Mean ± SEM	Mean ± SEM	Mean ± SEM	Mean ± SEM	Mean ± SEM	Mean ± SEM
D1	1.0±0.3	1.1±0.4	1.2±0.4	0.6±0.2	1.9±0.5	1.5±0.4	1.7±0.4	2.1±0.4
D15	1.0±0.3	2.5±0.6	1.2±0.4	0.6±0.2	1.9±0.5	1.5±0.4	1.7±0.4	2.0±0.5
D2	2.1±0.4	3.3±0.5	0.2±0.2	1.0±0.3	0.9±0.3	1.4±0.4	2.5±0.5	2.2±0.5
E1	2.0±0.4	2.3±0.5	0.7±0.3	0.6±0.3	1.8±0.4	2.2±0.5	1.5±0.4	2.3±0.4
G1	2.1±0.5	1.7±0.5	0.2±0.2	0.5±0.2	1.2±0.4	1.5±0.4	2.0±0.4	2.2±0.5
G15	2.1±0.5	1.6±0.4	0.2±0.2	0.4±0.2	1.2±0.4	2.2±0.5	2.0±0.4	1.1±0.3
G2	1.1±0.3	2.9±0.5	0.8±0.3	0.8±0.4	1.4±0.4	2.2±0.6	1.8±0.4	2.4±0.5
H1	2.1±0.5	2.2±0.7	0.7±0.3	0.8±0.3	1.4±0.4	2.3±0.6	2.0±0.5	2.3±0.5
I1	1.8±0.4	1.2±0.4	0.1±0.1	0.2±0.2	1.2±0.4	1.6±0.5	2.0±0.5	2.0±0.5
I5	2.2±0.5	2.5±0.5	0.1±0.1	0.4±0.2	0.8±0.3	0.9±0.4	1.5±0.4	1.4±0.4
K1	1.2±0.4	1.9±0.4	0.7±0.3	0.5±0.2	1.3±0.4	2.0±0.4	1.6±0.4	2.0±0.5
K2	2.2±0.5	2.5±0.5	0.5±0.2	0.5±0.2	1.7±0.5	1.6±0.4	1.8±0.4	2.0±0.5
Mean/Region	1.7±0.1	2.1±0.2	0.6±0.1	0.6±0.1	1.4±0.1	1.7±0.1	1.8±0.1	2.0±0.1

Heterogeneity is measured using Hamming distance and the results show the mean value per clone in each region for passage 1 and 11. The final row shows the heterogeneity average value of the passages in each region. Values between regions cannot be compared as we are expressing them as number of differences between clones, not taking into account the region's length, although this comparison can be made between passages.

doi:10.1371/journal.pone.0010319.t002

cells and are expressed as the 50% tissue culture infective dose per ml (TCID₅₀/ml). Virus was recovered from the culture supernatant when cytopathology was complete (5 to 7 days post-infection). For each subsequent passage, fresh MT-4 cells were infected with the same volume of the supernatant from the previous passage, taking into account that viral titers do not change significantly during serial passages in MT-4 cells [2]. p24 determination was carried out in the culture supernatant with an automated

enzymatic method in an Elecsys 2010 analyser using a p24 detection kit (Roche Diagnostics, HIV Ag.HIV (groups M and O) p24 antigen, ref.11971611 122). Cells were cultured in closed bottles and all the procedures were aimed at minimizing the possibility of cross-contamination, which was monitored with uninfected control cultures maintained in parallel. No evidence of contamination was observed at any time.

Fitness assay

Relative fitness values were determined by growth-competition experiments as previously described [44]. Briefly, the assay consists in the coinfection of MT-4 cells with a known number of TCID₅₀/ml of the virus to be tested and of a reference clone (J1). The genome of clone J1 contains a deletion of 45 nucleotides in the region of *env* gene that encodes the V1-V2 loop; this deletion is readily detectable in a GeneScan analysis. Coinfections, performed in duplicate or triplicate, were carried out for a maximum of 5 passages. The proportion of the two viruses in each passage was quantified by Gene Scan analysis. The fitness values per competitive transfer series were calculated from the slope of exponential plot as previously described [44]. The statistical significance of fitness differences between the initial and final populations was tested using the Graphpad software.

RNA extraction and PCR amplification for complete genome analysis

Consensus genomic nucleotide sequences for the entire HIV-1 genome were determined for the initial and final viral populations as previously described [4]. Sequences were derived from viral RNA in the supernatant of the cell cultures. RNA was extracted with a guanidinium isothiocyanate lysis buffer and glass milk as previously described [4]. RT-PCRs used for the determination of the consensus genome sequence were performed using at least 2,000 copies of viral RNA as starting template. The first amplification was carried out using the Access RT-PCR System

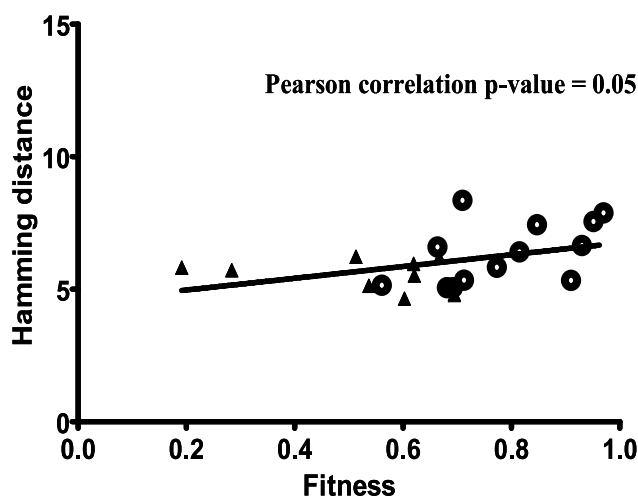


Figure 5. Correlation of heterogeneity and fitness in the HIV-1 clones during the serial passages. Fitness versus quasispecies heterogeneity values per clone are represented. Quasispecies heterogeneity was calculated as total Hamming distance, as explained in Materials and methods. A linear regression analysis between heterogeneity and fitness values was performed. The Pearson correlation test was used to check the values obtained, and the p-value is shown at the top. Symbols are as in Figure 2.

doi:10.1371/journal.pone.0010319.g005

(Promega) with 6 different pairs of oligonucleotide primers that cover the entire HIV-1 genome. Internal amplifications were performed with the EcoTaq DNA polymerase (Ecogen). The sequence of the primers used in the amplifications, covering the 5'LTR-gag (1880 bp), pol-vpr (3450 bp) and vpr-3'LTR (3923 bp) regions, will be made available upon request. Both external and internal amplifications involved 35 cycles with temperatures chosen according to the composition of the primers. Negative amplification controls were run in parallel to monitor absence of contamination. Primer location and numbering of nucleotide changes are referred to isolate HIV-1 HXB2 included in the Los Alamos data base [45]. Sequence was determined on the two cDNA strands with an ABI 373 automatic sequencer. Multiple sequence alignments were performed using the CLUSTAL W program [46]. Sequences have been deposited in GeneBank with the accession numbers GQ386774 to 386794.

GeneScan Analysis

For fitness determination, to avoid recombination during the reverse transcriptase reaction, quantification of the viruses was performed on proviral DNA. Prior to this decision, a comparison between estimations performed on proviral DNA or from viral RNA from the culture supernatant showed no statistical differences (data not shown). DNA was extracted from the cellular pellets of the competition cultures by standard phenol-chloroform method. The region encoding the V1/V2 loops of gp120 of the competing viruses, which were different in length, was amplified by PCR using one of the primers labelled with the fluorescent dye 6-FAM. The primers used were: 34, 5'-6FAM-GTCACAGTCTATTATGGGGTACCTGTGT-3'; and 31, 5'-ACCTCAGTCATTACACAGGCCTGCAGCGC-3'. The PCR products were diluted from 2-to 10-fold with water to obtain peak intensities in the linear range, and 1 µl of these dilutions were mixed with 9 µl of ROX 1000 size standard (Applied Biosystems) previously diluted 17-fold with formamide. Samples were run on an ABI PRISM 3700 machine, and data were analysed using

GeneMapper v 4.0. Peak areas were used to calculate the fitness of each virus versus the standard J1.

Quasispecies analysis and calculation of population heterogeneity

For the analysis of mutant spectra we selected four genomic regions that were amplified using three primers pairs. The first pair amplified RNA positions 672 to 1363 of HXB2 (692 nucleotides) that spanned the leader sequence and the p17 protein in the *gag* gene. The second pair amplified RNA positions 6045 to 6734 (711 nucleotides) encompassing *vpu* and the V1-V2 region of *env*. The third pair amplified RNA positions 7039 to 7641 (602 nucleotides) that comprised the V3-V4 region of *env*. In total, the three amplifications covered four regions: the 5' LTR/p17, *vpu*, *env* (V1-V2)- and *env* (V3-V4)-coding regions which represent a total of ~2,005 genomic residues. At least 20 independent clones were used to represent the viral quasispecies of each population at each passage. The RT-PCRs products were cloned using TA TOPO cloning (Invitrogen), following the manufacturer's instructions. In each region, average Hamming distance was calculated for every virus using the MEGA v4.0 software. A sum of the average Hamming distances in the 2,005 nucleotides analyzed was used for the analysis showed in Figure 4 and 5.

Statistical analysis of the results

Wilcoxon signed rank tests were applied to compare the different parameters used to characterize the viral populations. The p-values obtained are shown in each Figure, and the sum of signed ranks are indicated in the corresponding legends. A Pearson test of correlation was used to study the possible association between viral heterogeneity and fitness.

Author Contributions

Conceived and designed the experiments: AVB CLG. Performed the experiments: AVB RLR MP CC TA. Analyzed the data: AVB RLR MP CC TA ED CLG. Wrote the paper: AVB RLR ED CLG.

References

- Domingo E, Parrish CR, Holland JJ, eds. Origin and Evolution of Viruses (second Edition): Elsevier.
- Sanchez-Palomino S, Rojas JM, Martinez MA, Fenyő EM, Najera R, et al. (1993) Dilute passage promotes expression of genetic and phenotypic variants of human immunodeficiency virus type 1 in cell culture. *J Virol* 67(5): 2938–2943.
- Yuste E, Sanchez-Palomino S, Casado C, Domingo E, Lopez-Galindez C (1999) Drastic fitness loss in human immunodeficiency virus type 1 upon serial bottleneck events. *J Virol* 73(4): 2745–2751.
- Yuste E, Borderia AV, Domingo E, Lopez-Galindez C (2005) Few Mutations in the 5' Leader Region Mediate Fitness Recovery of Debilitated Human Immunodeficiency Type 1 Viruses. *J Virol* 79: 5421–5427.
- Eigen M (1971) Selforganization of matter and the evolution of biological macromolecules. *Naturwissenschaften* 58: 465–523.
- Domingo E, Martinez-Salas E, Sobrino F, de la Torre JC, Portela A, et al. (1985) The quasispecies (extremely heterogeneous) nature of viral RNA genome populations: biological relevance—a review. *Gene* 40: 1–8.
- Eigen MaBC (1988) Sequence space and quasispecies distribution. In: Domingo E, Holland JJ, Ahlquist P, eds. *RNA genetics*. Boca Raton: CRC Press. pp 211–245.
- Vignuzzi M, Stone JK, Arnold JJ, Cameron CE, Andino R (2006) Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population. *Nature* 439: 344–348.
- Pfeiffer JK, Kirkegaard K (2005) Increased fidelity reduces poliovirus fitness and virulence under selective pressure in mice. *PLoS Pathog* 1: e11.
- Jerzak GV, Bernard K, Kramer LD, Shi PY, Ebel GD (2007) The West Nile virus mutant spectrum is host-dependant and a determinant of mortality in mice. *Virology* 360: 469–476.
- Sauder CJ, Vandenburgh KM, Iskow RC, Malik T, Carbone KM, et al. (2006) Changes in mumps virus neurovirulence phenotype associated with quasispecies heterogeneity. *Virology* 350: 48–57.
- Vignuzzi M, Wendt E, Andino R (2008) Engineering attenuated virus vaccines by controlling replication fidelity. *Nat Med* 14: 154–161.
- Salazar-Gonzalez JF, Bailes E, Pham KT, Salazar MG, Guffey MB, et al. (2008) Deciphering Human Immunodeficiency Virus Type 1 Transmission and Early Envelope Diversification by Single-Genome Amplification and Sequencing. *J Virol*. pp 3952–3970.
- Escarmis C, Davila M, Domingo E (1999) Multiple molecular pathways for fitness recovery of an RNA virus debilitated by operation of Muller's ratchet. *J Mol Biol* 285: 495–505.
- Novella IS, Duarte EA, Elena SF, Moya A, Domingo E, et al. (1995) Exponential increases of RNA virus fitness during large population transmissions. *Proceedings of the National Academy of Sciences of the United States of America* 92: 5841–5844.
- Yuste E, Lopez-Galindez C, Domingo E (2000) Unusual distribution of mutations associated with serial bottleneck passages of human immunodeficiency virus type 1. *J Virol* 74: 9546–9552.
- Lopez-Galindez C, Lopez JA, Melero JA, de la Fuente L, Martinez C, et al. (1988) Analysis of genetic variability and mapping of point mutations in influenza virus by the RNase A mismatch cleavage method. *Proc Natl Acad Sci U S A* 85: 3522–3526.
- Thome KC, Radfar A, Rosenberg N (1997) Mutation of Tp53 contributes to the malignant phenotype of Abelson virus-transformed lymphoid cells. *J Virol*. pp 8149–8156.
- Larder B, Kemp S (1989) Multiple mutations in HIV-1 reverse transcriptase confer high level resistance to zidovudine (AZT). *Science* 246: 1155–1158.
- Diez J, Davila M, Escarmis C, Mateu MG, Dominguez J, et al. (1990) Unique amino acid substitutions in the capsid proteins of foot-and-mouth disease virus from a persistent infection in cellculture. *Proc Natl Acad Sci USA* 64: 5519–5528.
- Lopez-Galindez C, Rojas JM, Najera R, Richman DD, Perucho M (1991) Characterization of genetic variation and 3'-azido-3'-deoxythymidine- resistance mutations of human immunodeficiency virus by the RNase A mismatch cleavage method. *Proc Natl Acad Sci U S A* 88: 4280–4284.

22. Olivares I, Shaw G, Lopez-Galindez C (1997) Phenotypic switch in a spanish HIV Type 1 isolate on serial passage on MT-4 cells. *AIDS Res Hum Retroviruses* 13(11): 979–984.
23. Baranowski E, Sevilla N, Verdaguer N, Ruiz-Jarabo CM, Beck E, et al. (1998) Multiple virulence determinants of foot-and-mouth disease virus in cell culture. *J Virol* 72: 6362–6372.
24. Sanz-Ramos M, Diaz-San Segundo F, Escarmis C, Domingo E, Sevilla N (2008) Hidden Virulence Determinants in a Viral Quasispecies In Vivo. *J Virol*. pp 10465–10476.
25. Novella IS, Ebendick-Corpus BE (2004) Molecular basis of fitness loss and fitness recovery in vesicular stomatitis virus. *J Mol Biol* 342: 1423–1430.
26. Gonzalez-Lopez C, Gomez-Mariano G, Escarmis C, Domingo E (2005) Invariant aphthovirus consensus nucleotide sequence in the transition to error catastrophe. *Infect Genet Evol* 5: 366–374.
27. Domingo E, Gonzalez-Lopez C, Pariente N, Airaksinen A, Escarmis C (2005) Population dynamics of RNA viruses: the essential contribution of mutant spectra. *Arch Virol Suppl*. pp 59–71.
28. Shafer RW, Rhee SY, Pillay D, Miller V, Sandstrom P, et al. (2007) HIV-1 protease and reverse transcriptase mutations for drug resistance surveillance. *Aids* 21: 215–223.
29. Brander C, Walker BD (1999) T lymphocyte responses in HIV-1 infection: implications for vaccine development. *Curr Opin Immunol* 11: 451–459.
30. Borrow P, Lewicki H, Hahn BH, Shaw GM, Oldstone MBA (1994) Virus-specific CD8+cytotoxic T-lymphocyte activity associated with control of viremia in primary human immunodeficiency virus type 1 infection. *Journal of Virology* 68: 6103–6110.
31. Kaslow RA, Carrington M, Apple R, Park L, Mu-oz A, et al. (1996) Influence of combinations of human major histocompatibility complex genes on the course of HIV-1 infection. *Nature Medicine* 2: 405–411.
32. Cao J, McNeven J, Holte S, Fink L, Corey L, et al. (2003) Comprehensive analysis of human immunodeficiency virus type 1 (HIV-1)-specific gamma interferon-secreting CD8+T cells in primary HIV-1 infection. *J Virol* 77: 6867–6878.
33. Draenert R, Le Gall S, Pfafferoth KJ, Leslie AJ, Chetty P, et al. (2004) Immune selection for altered antigen processing leads to cytotoxic T lymphocyte escape in chronic HIV-1 infection. *J Exp Med* 199: 905–915.
34. Altfeld M, Addo MM, Shankarappa R, Lee PK, Allen TM, et al. (2003) Enhanced detection of human immunodeficiency virus type 1-specific T-cell responses to highly variable regions by using peptides based on autologous virus sequences. *J Virol* 77: 7330–7340.
35. Goonetilleke N, Liu MKP, Salazar-Gonzalez JF, Ferrari G, Giorgi E, et al. (2009) The first T cell response to transmitted/founder virus contributes to the control of acute viremia in HIV-1 infection. . pp jem.20090365.
36. Richman DD, Wrin T, Little SJ, Petropoulos CJ (2003) Rapid evolution of the neutralizing antibody response to HIV type 1 infection. *Proc Natl Acad Sci U S A* 100: 4144–4149.
37. Meyerhans A, Cheymier R, Albert J, Seth M, Kwok S, et al. (1989) Temporal fluctuations in HIV quasispecies in vivo are not reflected by sequential HIV isolations. *Cell* 58: 901–910.
38. Troyer RM, Collins KR, Abrahama A, Fraundorf E, Moore DM, et al. (2005) Changes in Human Immunodeficiency Virus Type 1 Fitness and Genetic Diversity during Disease Progression. *J Virol* 79: 9006–9018.
39. Wolinsky SM, Korber BT, Neumann AU, Daniels M, Kunstman KJ, et al. (1996) Adaptive evolution of human immunodeficiency virus-type 1 during the natural course of infection. *Science* 272: 537–542.
40. Shankarappa R, Margolick JB, Gange SJ, Rodrigo AG, Upchurch D, et al. (1999) Consistent viral evolutionary changes associated with the progression of human immunodeficiency virus type 1 infection. *J Virol* 73: 10489–10502.
41. Biebricher CK, Eigen M (2006) What is a quasispecies? *Curr Top Microbiol Immunol* 299: 1–31.
42. Eigen M, Biebricher CK (1988) Sequence space and quasispecies distribution; Domingo E, Holland JJ, Ahlquist P, eds. Boca RatonFlorida: CRC Press. pp 211–245.
43. Harada S, Koyanagi Y, Yamamoto N (1985) Infection of HTLV-III/LAV in HTLV-I-carrying cells MT-2 and MT-4 and application in a plaque assay. *Science* 229: 563–566.
44. Holland JJ, de la Torre JC, Clarke DK, Duarte E (1991) Quantitation of relative fitness and great adaptability of clonal populations of RNA viruses. *J Virol* 65(6): 2960–2967.
45. Kuiken C, Leitner T, Foley B, Hahn B, Marx P, et al. (2009) HIV Sequence Compendium 2009. Los Alamos National Laboratory, Theoretical Biology and Biophysics, Los Alamos New Mexico LA-UR 09-03280.
46. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22: 4673–4680.

Mutant spectra in virus behavior

Celia Perales^{1,2}, Ramón Lorenzo-Redondo³, Cecilio López-Galíndez³, Miguel Angel Martínez⁴ & Esteban Domingo^{1,2}

¹Centro de Biología Molecular Severo Ochoa (CSIC-UAM), C/ Nicolás Cabrera, 1 Universidad Autónoma de Madrid, Cantoblanco, Madrid 28049, Spain ²Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBERehd), Barcelona, Spain ³Centro Nacional de Microbiología (CNM), Instituto de Salud Carlos III, Majadahonda, Madrid 28220, Spain ⁴Fundació Irsi Caixa, Universitat Autònoma de Barcelona, Badalona, Spain

*Author for correspondence: Tel.: +91 196 4540 ■ Fax: +91 196 4420 ■ edomingo@cbm.uam.es

RNA viruses replicate as complex mutant spectra, also termed 'mutant clouds', known as viral quasispecies. While this is a widely observed viral population structure, it is less known that a number of biologically relevant features of this important group of viral pathogens depend on (or are strongly influenced by) the complexity and composition of mutant spectra. Among them, fitness increase or decrease depending on intrapopulation complementation or interference, selection triggered by memory genomes, pathogenic potential of viruses, disease evolution and the response to antiviral treatments. Quasispecies represent the recognition of complex behavior in viruses, and it is an oversimplification to equate such a population structure with the classic polymorphism of population biology. Darwinian principles acting on genome collectivities that replicate with high error rates provide a unique population structure prone to flexible and largely unpredictable behavior.

Viral quasispecies & their relevance for viral population dynamics

RNA viruses in nature form complex distributions of closely related but nonidentical genomes subjected to a continuous process of genetic variation, competition and selection. Such mutant distributions are termed 'viral quasispecies', and their existence *in vivo* has been documented by analysis of molecular and biological clones isolated from viral populations, as well as by the application of the new ultradeep sequencing techniques. This means that, counterintuitively, no defined genomic sequence characterizes a viral population at any given time, except as an average of many different sequences. This population structure was first documented with bacteriophage Q β during replication in its *Escherichia coli* host [1–3], and was later evidenced for many RNA viruses, including important human and animal pathogens, such as HIV-1, human hepatitis C virus (HCV), vesicular stomatitis virus (VSV) and foot-and-mouth disease virus (FMDV), reviewed in [4–9]. Virologists adopted the term quasispecies from the name used by Eigen and Schuster to theoretically describe the type of population structure proposed to have mediated self-organization and adaptability of primitive replicons at the early stages of life development on Earth [10,11]. The connection between quasispecies theory and the population dynamics of viruses was made at a Max Planck seminar in 1978 (accounts in [12–15]). Ever since, viruses have been regarded by theoretical biologists as

experimental model systems for quasispecies dynamics, and virologists have found support to interpret many experimental observations both in infections in cell culture and *in vivo* from quasispecies [6,12,13,16,17].

Discrepancies regarding the adequacy of the term quasispecies to describe the evolutionary dynamics of viruses have been expressed by some population biologists, although this took more than two decades to be manifested (see [18], the exchange of letters [19,20] or the reaction of [21] to [15]). The main claim, as we understand it, is that for the term quasispecies to be used correctly, the type of variation observed for viruses should correspond to steady-state mutant distributions that have attained a mutation-selection balance. However, viral quasispecies will be perturbed frequently by environmental fluctuations and, therefore, transient equilibria (constancy of a consensus genomic sequence with an underlying dynamic mutant spectrum) may be difficult to observe. Viral populations under a transient equilibrium or away from equilibrium have a population structure of a highly dynamic swarm of mutants, rather than defined mixtures of variants (as in classical polymorphism). Nevertheless, as covered in the following sections of this article, several examples of equilibrium (or near-equilibrium) conditions in viral populations, as well as of deterministic behavior, have been described. Where this is not sufficient to refer to RNA viral populations as quasispecies, extensions of quasispecies theory to complex fitness landscapes (that more realistically apply

Keywords

- antiviral treatment
- complementation
- Darwinian principles ■ error rate ■ interference ■ mutant spectrum ■ quasispecies
- RNA virus ■ vaccine approaches ■ viral fitness
- viral pathogenesis

future medicine part of fsg

to RNA viruses) have been developed [22–24]. It is common procedure in theoretical biology to formulate a deterministic model that is mathematically approachable, and then to introduce stochastic effects that render the model more realistic. The extensions of the initial deterministic quasispecies theory that were articulated to include stochastic components are an example. The initial quasispecies theory may not be realistic, even for the primitive replicons it intended to represent.

The reluctance of some population geneticists to generalize the use of the term quasispecies may have been influenced by the inheritance of a belief in a limited weight of mutations as promoters of evolutionary change, championed by some important and influential schools of population geneticists. The reader will find it illustrative to compare the views expressed by Wright [25] on the impact of mutation on general evolution:

“There must be gene mutation, but an excessive rate gives an array of freaks, not evolution.”

and by Holland and colleagues on mutation as a hallmark of RNA genetics [7]:

“This high mutation frequency is coupled with high rates of replication reflected in rates of RNA genome evolution, which can be more than a million fold greater than the rates of the DNA chromosome evolution of their hosts.”

Even recently, high mutation rates have, on occasion, been regarded as being a consequence of rapid RNA virus replication rather than being an adaptive trait for RNA viruses [26]. In fact, the main departure of quasispecies from other mathematical treatments of evolutionary and ecological dynamics is precisely the prominent role of mutations [27].

Virologists use the term quasispecies to mean ‘distributions of variant genomes subjected to a continuous process of genetic variation, competition and selection, and that can act as a unit of selection’ [5]. Genetic variation involves mutation and recombination [28]. This use by virologists implies some modest modifications of the general definition of quasispecies given by Nowak [29,30]:

“Quasispecies is an ensemble of similar genomic sequences generated by a mutation-selection process.”

In the course of RNA viral infections *in vivo*, multiple mutant distributions may arise in different cells, tissues or organs of the same infected host individual. They may have arisen by mutation and selection, or following a random drift event by which one or few founder individuals reach a cell, tissue or organ. Variant viruses may also be the result

of multiple independent infections by related virus variants, or of reactivation of persistent or latent forms whose replication may be triggered or enhanced as a result of reinfection ([31–33]; review in [34]). Phylogenetic analyses may allow the distinguishing of independent variants (i.e., resulting from reinfection or reactivation) from closely related variants that derive from an initial genome by mutation. However, at times, the distinction may not be statistically solid, and it may not be possible to ascertain whether competing variants were all derived from the same viral genome. If the different variants are related and participate in replicative competition, we do not see a powerful conceptual barrier to considering them as part of the ‘viral quasispecies’. These are the types of extensions that even population geneticists have accepted when applying principles, which, in their original formulation, were unrealistic [6,19]. Competition among different variants of a virus can be established at different levels: within single infected cells; within a population of cells; or in a broader host compartment that has received multiple swarms from individual quasispecies, such as in blood during viremia, where related variants may meet and compete to enter some cell types or to evade immune responses (i.e., circulating neutralizing antibodies in serum).

In all cases in which an infection has been initiated by a single infectious particle or by a clonal population, either in cell culture or in pathogen-free animals or plants under controlled conditions, it has been shown that multiple variants arise to produce complex mutant spectra (as an early example of an experiment *in vivo*, see [35] and the comment by Temin relating HIV-1 and FMDV population dynamics *in vivo* [36]; for plant viruses see [37]). Statistical evaluation of the viral genomic sequences in such controlled experiments has indicated that the number of variants that arise is not limited to a few ‘alleles’ subjected to balanced selection (as in most accepted models to explain classical polymorphisms in population genetics). Rather, an ample and largely indeterminate swarm of mutants often referred to as ‘mutant spectrum’ or ‘mutant cloud’ is generated, whose composition varies as a function of time. Sequential dominance of genome subpopulations was documented in early studies on persistent viral infections *in vivo* [35,38,39]. Transient equilibrium distributions are expected to be perturbed by internal factors (dependent on the mutant distributions themselves) and external factors (environmental parameters). As surprising and counterintuitive as it might be, the fact

is, with a high probability, each existing mutant distribution of any virus has a unique composition regarding the minority genomes that compose the population. Ultradeep sequencing procedures are currently contributing to quantifying the complexity of viral populations (by determining thousands of nucleotide sequences per sample rather than tens by molecular cloning and Sanger sequencing), and to unveil differences in mutant spectrum composition among sequential virus samples from the same infected individual. That is, these procedures have confirmed the occurrence of mutation and competition *in vivo*, involving a multitude of minority sequences. The reason why we resort to writing a consensus sequence is because of necessity and tradition. A sequence reassuringly associates the virus with a given genomic organization, a set of regulatory signals and a number of encoded proteins of a defined amino acid sequence. No concise means to denote a mutant spectrum has been developed yet, although possible advantages of a new-generation data bank within the current virogenomics procedures that could include information on minority mutants have been suggested [40].

The main contribution of the quasispecies theory to the understanding of virus behavior at the population level has been to explain the adaptive capacity of viruses through their population structure. The fact that viral populations consist of complex and dynamic mutant distributions impels their capacity to occupy sequence space in a way that is unparalleled by classical polymorphic sites [12,41]. Sequence space is a theoretical representation of all possible variants of a nucleotide sequence [41]. It is as if any mutant swarm were pluripotent, in that it can contribute subsets of individuals (or of groups of individuals) able to respond to a selective constraint that limits the replication of the majority of the population. Being a mutant spectrum rather than a combination of defined types underlies the major impact of the quasispecies population structure. The difference is not merely quantitative. It has forced revision of how to deal with the genotype–phenotype relationship in viruses, and has influenced preventive and therapeutic control strategies to combat viral disease (see also section on Mutant spectra and polymorphic loci: two views of viruses as populations) [6,12,15,42–44].

The present article reviews experimental evidence of quasispecies that illustrates how critical it is to view viruses as mutant spectra, and how many basic features of viruses can be

missed if we only consider consensus sequences. We include examples of steady-state mutant distributions, and describe the evidence in favor of ensembles of mutants as the unit of selection, and the ‘advantage of the flattest’ effect (the fact that a viral population located on a flat fitness surface can be more resistant to the deleterious effect of mutations than a virus that occupies a high fitness peak, a prediction made with digital organisms). We emphasize that ensembles of mutants are selected when viral quasispecies are confronted with selective constraints, and that even individual viral clones can self-organize to produce a phenotype that can be understood only as a result of intramutant spectrum interactions. These are not ‘mystical’ properties of quasispecies, but properties that result from intrapopulation molecular interactions during viral replication, which have begun to be recognized only recently [45–48]. Collective behaviors typical of viral quasispecies have been observed with cell collectivities and prions, thus extending the scope of quasispecies-related concepts and opening new avenues of therapeutic interventions for some nonviral infectious and proliferative diseases [15,49]. Finally, the differences in antiviral approaches prior to, and after, the quasispecies concept came into play, are commented on as a demonstration of the practical impact of the introduction of quasispecies in virology.

Presence of mutant clouds is supported by biochemical data

The breadth of the composition of the mutant spectrum of RNA viruses originates from the stochastic occurrence of mutations, and from population disequilibria, which result in a change of the cloud composition with time. Poliovirus RNA replication constitutes a revealing example. It takes less than 1 min to synthesize one full-length poliovirus RNA chain inside an infected cell, and 2000–3000 RNA molecules can be produced per cell [50]. Several RNA molecules can be synthesized simultaneously in membrane-associated replication complexes, and the entire infectious cycle can be completed in a few hours [51,52]. Measurements of the mutation rate or frequency for poliovirus have yielded values that range from fewer than 10^{-6} substitutions per nucleotide (s/nt) based on repetitive nucleotide sequencing [53], 10^{-5} – 10^{-4} s/nt based on the frequency of mutants resistant to monoclonal antibodies [54–56], 2×10^{-6} s/nt from reversion of an amber mutation [57], 2×10^{-5} s/nt from reversion of a point mutation [58], and a range of 2×10^{-3} to 4×10^{-3} s/nt by analysis of

RNase T1-resistance (determining the proportion of minority nucleotides at RNA sites dominated by G residues) at defined genomic positions [59], a method developed by Steinhauer and Holland [60]. In these comparisons, no distinction was made between values that may be close to a true mutation rate (the proportion of misincorporation events during viral RNA synthesis) or that represent a mutation frequency (the proportion of mutants or a mutant type in a population). This important distinction, the confusion in the literature deriving from the frequent use of mutation rate to mean the rate of accumulation of mutations with time (or rate of evolution), and the effects of fitness on determining mutation rates and frequencies, have been discussed previously [4,5,61].

Mutation rates can vary, not only with environmental parameters (intracellular ionic and metabolic conditions), but also with the sequence context of the template RNA, the amino acid sequence of the RNA-dependent RNA polymerase (RdRp), and under the influence of viral and host proteins [62–68]. In the case of poliovirus, a single amino acid substitution (G64S) in its RdRp was sufficient to decrease, by approximately threefold, the error frequencies in a guanidine-resistance assay, decrease the amplitude of the mutant spectrum sixfold, and modify the biological behavior of the virus ([48,69]; see also the section on Mutant spectrum complexity and viral pathogenesis). Since poliovirus forms mutant spectra that may include minority components with altered copying fidelity, actual mutation rates may be relatively flexible upon poliovirus replication, even when infection is initiated by a single genome or a clonal viral population, coherent with evolvability being a selectable trait [70]. For the main purpose of this article, a conservative estimate of the average poliovirus point mutation rate can be considered 10^{-5} s/nt (insertions and deletions are less frequent, except at homopolymeric tracts). Taking 7500 as the number of genomic nucleotides, the occurrence of one point mutation will require the synthesis of 13 polioviral RNA molecules. This means that, in each infected cell, 154–231 genomes with a mutation will be generated. Even assuming that 80% of the mutations do not produce RNA that can contribute significantly to the viral population heterogeneity (either because the mutations cause a severe fitness decrease or because they are lethal [71]), it means that ten infected cells can give rise to 308–462 mutants. Even if the number of susceptible cells in the oropharyngeal and intestinal mucosa in which

primary poliovirus replication takes place is not known, it can be anticipated that, in most infections, waves of mutant generation in infected cells will take place [51,72]. This description represents just a snapshot of a continuously evolving, highly dynamic process, in which in any newly infected cells, new rounds of mutants are produced and tested for survival capacity in their physical and biological environments. The figures used for the aforementioned calculations are likely to lead to an underestimate of the real population heterogeneity. Thus, both the rate of RNA synthesis and the mutation frequencies strongly support the notion of rapid generation of mutant clouds. No evidence exists to suggest that the course of events could be significantly different for other RNA viruses [73], despite viruses varying greatly in the number of particles produced by infected cells [74]. Not surprisingly, quasispecies has become a fundamental concept in clinical virology, with multiple implications for viral disease prevention and treatment [15,42,43,75].

An independent line of biochemical research that supports the formation of mutant clouds is based on measurements of copying fidelity of viral polymerases *in vitro*. With rare exceptions, both viral reverse transcriptases (RTs) and RdRps lack a domain corresponding to a 3'–5' exonuclease repair function, equivalent to that found in replicative cellular DNA-dependent DNA polymerases [76,77]. Biochemical assays also failed to provide evidence of proofreading exonuclease activity in the virion-associated RdRp of VSV [78]. This appears to be the rule, despite repair functions having been described for some RNA genomes (reviewed in [73]), and evidence that an exonuclease function is operational in some coronaviruses [79,80]. Measurements of template-copying fidelity with RTs and RdRps *in vitro* have yielded a wide range of values, depending on the assay employed, the type of misincorporation measured, nucleotide sequence context of the template primer (that may determine whether direct misincorporation or misalignment mutagenesis prevails), presence of additional viral or host proteins, relative concentration of nucleotide substrates, ionic conditions in the reaction mixture, and other influences [4,81–83]. The values obtained with purified enzymes are very often in the range of 10^{-3} – 10^{-5} misincorporations per nucleotide, consistent with the great potential to generate mutant genomes during replication of the corresponding viruses in infected cells. The mutation rate of HIV-1 measured *in vivo* is approximately

20-fold lower than predicted from the fidelity measured with purified RT [84]. When the proof-reading exonuclease activity of replicative DNA polymerases is absent or suppressed, misincorporation rates approach those observed with RTs and RdRps [85]. Thus, for viruses that replicate actively in their infected hosts, biochemical data on the enzymology of viral genome replication is consistent with the observations of nucleotide sequence heterogeneity measured for evolving viral populations. To our knowledge, no evidence of genetic homogeneity has been obtained for any replicating RNA virus, even in clonal populations.

In addition to mutations made by viral polymerases, there are other mechanisms implicated in the generation of mutant clouds. HIV recombines at exceedingly high rates [86], approximately one order of magnitude more frequently than simple gammaretroviruses, such as murine leukemia virus and spleen necrosis virus. A precise calculation of the HIV-1 recombination rate has rendered an average result of approximately three crossovers per genome per round of replication, with a range of one to seven crossovers [87]. Thus, the recombination rate for HIV-1 is approximately tenfold greater than the point substitution rate (~ 0.25 per genome per round) [84]. Given the dynamics of HIV-1 turnover *in vivo*, and a recombination rate of approximately three crossovers per cycle, some genome lineages from a 15-year-old infection may have experienced as many crossovers as base mutations in the genome. It has been proposed that recombination coupled with mutation profoundly influences HIV evolution, and gives it a nonclonal and transient nature *in vivo* [88]. An example of the adaptive potential of HIV-1 recombination is the fact that multidrug-resistant variants of HIV-1 can exist in cells as defective quasispecies, and can be rescued by superinfection with other defective HIV-1 variants [89]. This result is most likely attributable to recombination during second rounds of infection, and suggests that defective HIV-1 variants may, nonetheless, constitute part of the HIV-1 reservoir [90]. Host cell ssDNA cytidine deaminases (APOBEC3) can extinguish HIV-1 infectivity by its incorporation into virus particles and subsequent cytosine deaminase activity that attacks the nascent viral cDNA during reverse transcription, causing lethal mutagenesis. Recently, it has been demonstrated that APOBEC3G can also induce sublethal mutagenesis, which would maintain virus infectivity and contribute to HIV-1 variation [91]. Mutation by host cell APOBEC3 deaminases is not restricted to retroviruses. Hepadnaviruses,

such as hepatitis B virus (HBV), are also vulnerable to mutation by APOBEC3 [92]. Although the mutant spectrum resulting from APOBEC3 editing is highly deleterious, a small fraction of lightly APOBEC3G-edited genomes can impact HBV replication *in vivo* and, possibly, immune escape [93]. Finally, both RT and cellular RNA Pol II may be influenced by cellular nucleotide concentrations [94], and may contribute to the overall virus genetic variation. These alternative sources for generating mutant clouds have the potential to play a role in viral evolution, pathogenesis, immune escape and drug resistance. Thus, HIV-1 diversity in infected individuals is increased by the compartmentalization of the infection, the presence of viral reservoirs and the reactivation of latent provirus. Different body compartments, different sites in the same organ, or even an infected cell, may harbor mutant and recombinant HIV-1 [86,95–98].

Mutant spectra & polymorphic loci: two views of viruses as populations

The differences between the classical concept of polymorphism of population genetics and a mutant spectrum in a viral population have been mentioned in the introductory section, and were discussed previously [12]. In this section, we add that the difference has some parallel with the distinction between the ‘classical’ and ‘balance’ hypotheses of genetic structure, that were debated in the early 20th Century (reviewed in [99]). The classical model proposed that, at each gene locus, the gene pool of a population is dominated by a wild-type, the frequency of which approaches 1. This model was utilized by Muller and some of his associates [99]. By contrast, the balance model proposed that most loci of a gene pool of a population consisted in a number of alleles present at moderate frequency. This results in an array of genotypes, which provide sufficient fitness in most environments that the population has to face. This model was utilized by Dobzhansky, Ford and their colleagues [99]. A similar conceptual distance to that between the classical and balance hypotheses exists between the balance and the mutant spectra of viral populations, as revealed by the experimental determinations of population heterogeneity.

The biological implications of a difference between a defined number of alleles and a mutant spectrum (FIGURE 1) are rendered even more dramatic by the fact that the former generally refers to complex genomes (typically cellular genomes with chromosomal DNA of 10^6 – 10^9 bp), while the latter refers to viral RNA genomes (typically

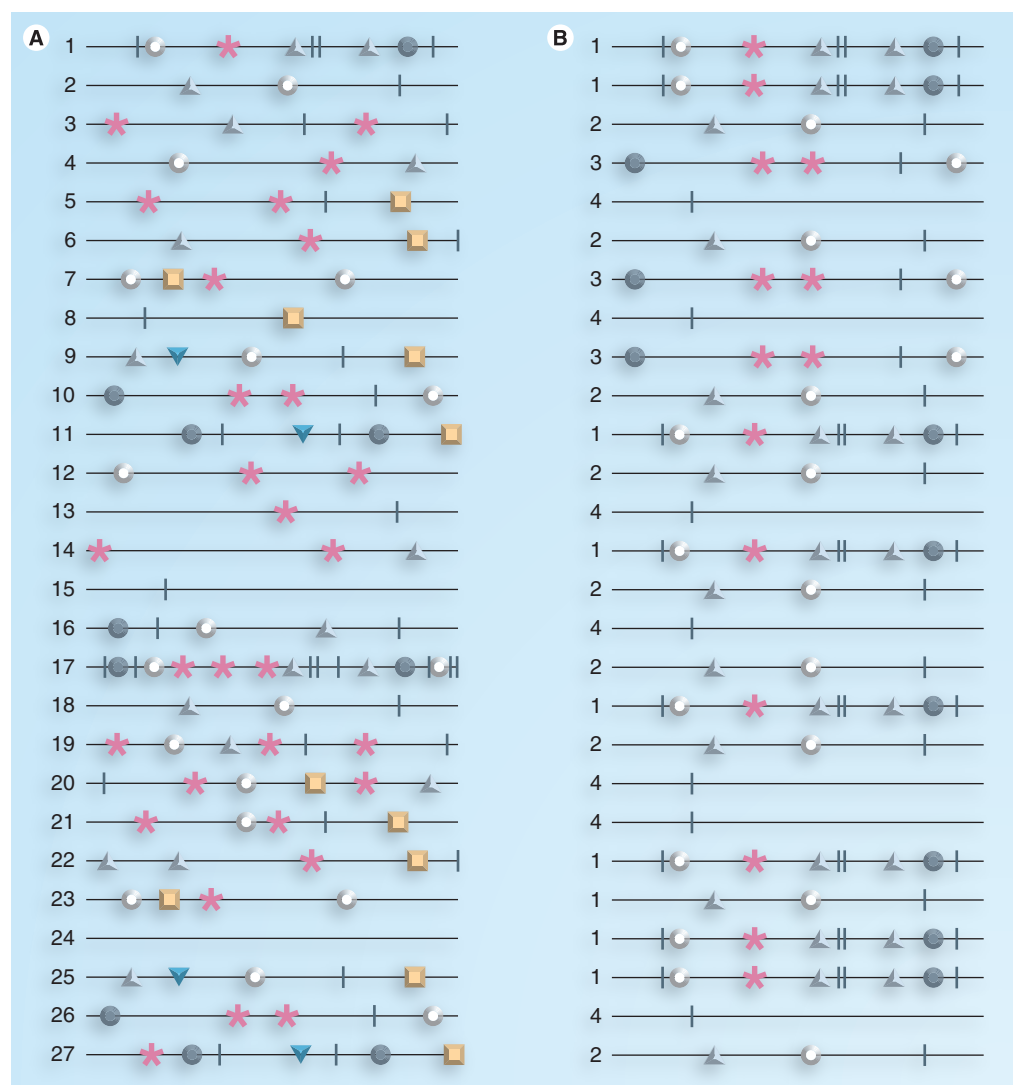


Figure 1. Two classes of mutant distributions. Genomes are depicted as horizontal lines and mutations are represented by different symbols on the lines. Despite both distributions looking similar, they are very different regarding occupation of sequence space. In distribution in (A), each genome includes a different set of mutations. By contrast, the distribution in (B) is made of only four types of genomes. In viral quasispecies (A), highly dynamic mutant spectra provide ample opportunity for phenotypic change. Identical mutations are often observed in viral quasispecies when a limited genomic region is compared, but identical full-length genomic sequences are rare, except in populations that are very close to a clonal origin. In polymorphic populations (B), the usually limited mutation rates provide little opportunity for phenotypic change. Scheme is based on experimental data discussed in the text.

of 3×10^4 to 3.3×10^5 nucleotides). The outcome is a vastly superior capacity of a viral mutant spectrum to explore sequence space compared with a polymorphic cellular organism. The theoretically possible number of single mutants of a 10,000 nucleotide genomic RNA sequence is 3×10^4 , while the population size of a virus can attain values that are several orders of magnitude larger. This means that all single mutants and many double mutants (as well as decreasing proportions of, for example, triple or quadruple mutants) are potentially present in most evolving

RNA virus populations. By contrast, for a typical mammalian genome, the potential number of single mutants will be approximately 10^{10} , a number that is well above the population size of any mammalian species. That is, even if several polymorphic loci are present in a population of differentiated organisms, the capacity to explore sequence space (and its corresponding phenotypic space) is tremendously inferior to the capacity displayed by viral replicons [12,41,100]. Furthermore, in classical population genetics, the definition of polymorphic loci excluded

alleles found only once (see, for example, [101]). The conceptual difference between polymorphic loci and mutant spectra (irrespective of the origin of the components of a given mutant spectrum) is clearly expressed in a passage of the Spiess book, which referred to the number of polymorphic alleles at a locus of *Drosophila pseudoobscura* [101]:

“Many alleles occurred only once in a single strain; they were suspected as being due to mutation and not falling within the definition of polymorphism (if an allele was found only once in a just a single locality, it was not counted toward polymorphism).”

Obviously, relaxation in the strict use of polymorphism is also justified. Historically, developments in biochemistry have forced reconsideration of the levels of diversity in natural populations of differentiated organisms. An example is the ‘zymogram’ technique, consisting in a combination of electrophoresis with *in situ* staining methods [102,103]. This procedure revealed the presence of allozymes (variants of the isozymes of esterases, phosphatases, phosphoglucosmutases and several dehydrogenases, for example, as a result of gene mutation) in humans [104] and *Drosophila pseudoobscura* [105,106], and indicated higher levels than anticipated of variation among individuals of the same species. A similar situation is currently witnessed as a result of the application of large-scale microarray and ultradeep sequencing methods, and metagenomic analyses [80,107–118]. Even if we relax the scope of application of ‘polymorphism’, a viral mutant spectrum can contain multitudes of mutants that will be scored only once, and perhaps never (or rarely) found again in other mutant distributions. Incidentally, these new sequencing techniques, together with bioinformatics-based structural analyses, may lead to the possibility of prediction of some changes related to antigenic variation (compare [113] with [119] and [120]). Thus, it might be possible to achieve the old dream of S Fazekas de St Groth of anticipating antigenic drift for vaccine design (see [121]; reviewed in [122]).

Thus, as illustrated in FIGURE 1, the major conceptual departure of quasispecies diversity versus multiple-allele diversity is that the former implies fleeting and broad collections of genomes with a rather indeterminate composition, whose behavior cannot be precisely assigned to a genomic sequence. The population structure of RNA viruses has sometimes been described with alternative terms, such as ‘polyploidy’, ‘swarms’, ‘heterospecies’, ‘heteropopulations’ and ‘clone’ [13,36]. Given the close similarities between the dynamics displayed by viral populations and the dynamics conveyed by

quasispecies theory, it did not seem justified to prefer an alternative term to ‘viral quasispecies’. To endow quasispecies with its more general (and experimentally operative) meaning, it may be helpful to stress that quasispecies have a physical, chemical and biological definition [13]. Physically, quasispecies are defined as a cloud in sequence space [41]. Chemically, quasispecies are a rated distribution of related but nonidentical genomic sequences, which is the definition familiar to virologists. Biologically, quasispecies are the target of selection. It is also increasingly appreciated by virologists that ‘ensembles’ of genomes, rather than ‘individual’ genomes are selected. Again, several meanings of ‘ensemble as a unit of selection’ can be envisaged. It can mean that, since individual RNA genomes have a fleeting existence (not the case of polymorphic sites), and their evolution as individual genomes is heavily influenced by the mutant spectrum surrounding them [123–125], unavoidably, a group of individuals must be selected. In reconstructed quasispecies, swarms of related antigenic variants were selected by a monoclonal antibody [126]. However, direct demonstration that an ensemble (or set of interacting individuals) is actually being selected, not only because they share a selectable trait, but because of their interactions, has also been demanded as a condition to use the term quasispecies. As described in the following paragraphs, no matter what the requirements imposed upon the terms ‘ensemble as a unit of selection’, there is experimental evidence of it being followed by viral populations. The need to focus on mutant spectra to understand viruses as collectivities was illustrated with an analogy in physics [13,127]:

“In virology, it is no more meaningful to discuss the environmental fitness of an individual RNA virus genome sequence than it is in statistical mechanics to refer to the temperature of a single gas molecule”.

Since the time that this comparison was made, overwhelming evidence has supported a biological role for interactions established within mutant spectra.

Complementation, interference & the non-neutrality of mutant distributions Mutant spectra as determinants of virus fitness & extinction during lethal mutagenesis

The behavior of viral quasispecies can be influenced by the complexity (number of different genomic sequences and the average number of mutations per genome) and composition (the types of sequences) present in the mutant

spectrum. In the following sections, we review a number of biological properties of viruses replicating either under standard (basal) mutation rates or elevated mutation rates (promoted by mutagenic agents), which can be interpreted only if the mutant spectrum is taken into consideration.

The first direct evidence that a mutant spectrum could influence the behavior of a specific mutant was obtained by de la Torre and Holland, who showed that a VSV quasispecies could suppress individual viral genomes of higher fitness [124], an observation in line with theoretical predictions [41,125]. In the course of lethal mutagenesis experiments (or virus transition into error catastrophe) of FMDV, it was observed that mutagenized FMDV populations on their way to extinction (pre-extinction populations) interfered with the replication of standard FMDV [128]. It was proposed that altered forms of viral proteins, produced as a consequence of the introduction of multiple mutations by the mutagenic agent, could act *in trans*, and perturb viral replication functions, a possibility that is compatible with the types of intermolecular interactions known to operate during viral replication [129]. *Trans*-acting mutants of poliovirus were shown to delay replication of drug-resistant poliovirus mutants [130]. Upon mutagenesis of the arenavirus lymphocytic choriomeningitis virus during a persistent infection of BHK-21 cells, loss of infectivity preceded the arrest of viral RNA replication. This suggested that noninfectious but replication-competent lymphocytic choriomeningitis virus accumulated as a result of mutagenesis, prior to virus extinction [131]. Such defective RNAs were termed 'defectors'. An *in silico* experiment on viral replication in the presence or absence of defectors suggested that the latter could be an important determinant of viral extinction. The generation and effect of defectors was predicted to be more prominent under low or moderate increases of the average mutation rate. Higher mutation rates would then be associated with overt lethality of viral genomes. These experimental and *in silico* observations led to the proposal of the lethal defection model of virus extinction [131], a model that has been further supported by theoretical studies [132]. The assumption that defector genomes may interfere with replication of standard infections of FMDV was confirmed experimentally in coinfection experiments, in which standard FMDV RNA was co-electroporated into BHK-21 cells, together with mutant RNAs. A number of polymerase and capsid FMDV mutants that had been previously characterized [133,134] were

tested for their interfering activity. Interference was observed with mutants that were competent in RNA replication, but not with mutants defective in RNA replication, and it did not require higher than basal mutation rates. Interestingly, interference occurred at later times after infection, while the same mutants produced an enhancement of standard FMDV replication early after infection, probably associated with transient complementation, by mechanisms that are not well understood [135].

A sustained basal level of complementation may be occurring normally in viral quasispecies. The evidence is provided by the observation that the relative fitness (replication capacity relative to a reference virus in a given environment) was significantly higher for an average Q β or VSV population than for individual biological clones sampled at random from the same population [3,136]. That a basal complementation may be normal is also consistent with the high proportion of deleterious mutations that are generated and maintained in viral quasispecies when they are well adapted to the environment in which replication takes place [71].

Complementation or interference among components of a mutant spectrum may play a role in virus behaviour, not only during the transition towards extinction but also under standard replication conditions, sometimes with medical implications [137,138]. Suppressive effects of mutant spectra [124] were observed in FMDV that had been subjected to multiple passages in cell culture in the presence of polyclonal antibodies [139]. When individual antigenic variants of FMDV were removed from their corresponding mutant spectra, their capacity to produce infectious progeny was increased 10⁴- to 10⁵-fold [139]. Complementation may sustain the presence of debilitated mutants in viral populations [140,141] and may permit a specific mutant to reach a target organ by virtue of being surrounded by an adequate mutant spectrum [48].

Despite being prone to shifts of viral subpopulations, some viruses may replicate extensively without modification of the genomic consensus nucleotide sequence. This appeared to be the case for the bacteriophage Q β when it was propagated in its natural host *E. coli* [3]. In this case, invariance was probably favored by a long history of adaptation of Q β to its host before the laboratory passages were initiated. This suggests a steady-state equilibrium population, despite continuous replication (a situation that, hopefully, few would deny deserves being referring to as quasispecies). Initial fitness recovery of HIV-1 clones that had

been debilitated by serial bottleneck passages did not necessitate a modification of the consensus sequence [142]. In a totally different environment – the one provided by mutagenic agents – constancy of the consensus sequence was also observed. A virus can be pushed towards extinction while undergoing logarithmic decreases in specific infectivity, with no effect on the consensus sequence [131,143]. Thus, not only under conditions of population equilibrium, but also when undergoing genetic deterioration without possibility of positive selection, consensus viral sequences may remain invariant. It should not escape the attention of the reader that an invariant consensus sequence does not mean absence of mutations, since a dynamic swarm lies behind – it mutates continuously to yield the same ‘average’ nucleotide sequence. As a consequence, highly relevant processes may go unnoticed if only consensus sequences (the ones that have been the main focus of classical genetics) are followed.

A replicating clonal population is also capable of self-organization in a constant environment. Passage of a clone of FMDV in cell culture resulted in its diversification into two genetically and phenotypically distinct subpopulations [45,46]. The two subpopulations, whose individual components were sampled and characterized, were termed ‘competitors’ and ‘colonizers’. Colonizers alone killed the host cells more efficiently than the competitors. In cells coinfecting with the two virus classes, viral interference by the competitors resulted in reduced cell killing, and competitors replaced colonizers, following dynamics previously recognized in classical ecological systems. A mathematical model predicted a density-dependent selection, which was confirmed experimentally [45,46]. Modulation of virulence through interactions among components of a viral quasispecies that derive from a single infectious genome documents the capacity of viral quasispecies to self-organize so as to express a phenotype beyond that of its individual components. Furthermore, the events that underlie modulation of virulence in this cell culture system may explain differences of virulence (that could not be attributed to differences in the consensus sequence or mutant spectrum complexity) affected by the frequency of coinfections *in vivo* [144].

Codon usage is an adaptive feature of viruses, which sometimes select rare codons to slow the rate of ribosomes along the viral RNA to allow proper protein folding, in a process termed fine-tuning translation kinetics selection [145]. This type of selection is displayed by hepatitis A virus

(HAV) to ensure the correct folding of the capsid proteins needed for virion stability, an essential feature for the extracellular survival of HAV and its transmission. In a study of readaptation of HAV to an environment with altered tRNA pools, selection for fine-tuning translation kinetics acted on the entire HAV mutant spectrum, rather than on individual genomes [146]. Constellations of mutations scattered among components of the mutant swarm were associated with fitness recovery. The results also showed faster adaptation of the HAV population during processes that took the population towards an environment that had been previously experienced by the same population [146]. This observation is consistent with the presence of a molecular memory in the HAV quasispecies, another feature that emphasizes mutant ensembles as the target of selection, as reviewed in the following section.

Memory genomes & their influence in evolution reveal deterministic features of virus behavior

As documented in previous sections, viral quasispecies can be regarded as a repository of genetic and phenotypic variants that can become dominant in response to environmental changes. Mobilization of minority elements as a reaction to an external stimulus is typical of complex adaptive systems, for example, the immune system of vertebrates [147,148]. When stimulated by a foreign antigen, the complex collectivity of cells from the immune system respond by undergoing a cascade of genetic changes that results in selection and clonal expansion of cells showing the highest affinity for the antigen. The process generates stable memory T cells that may last through the lifetime of the organism. Inspired by the behavior of the immune system as a complex adaptive system, we designed experiments with FMDV that demonstrated that genomes hidden in the mutant spectrum constituted a molecular record of the evolutionary history of the population [149]. Such a history-dependent record was termed quasispecies memory, and it was tested with two genetic markers of FMDV: an antigenic variant selected with a monoclonal antibody, and a variant with an elongated internal oligoadenylate that had become dominant upon subjecting FMDV to plaque-to-plaque (serial bottleneck) transfers. In both cases, when the selective pressure was withdrawn or the passage regime changed, the populations reverted to the wild-type quasispecies. The revertant

populations maintained the original variants at levels 10^2 - to 10^3 -fold higher than the levels that could be attributed to basal mutational pressure [149,150].

Investigations on quasispecies memory have been instrumental to document both the behavior of genome collectivities as the unit of selection, and deterministic features (as predicted by the initial quasispecies theory). The main evidence and implications are as follows. Quasispecies memory was erased by subjecting the population to a bottleneck event. This implies that memory is a property of the quasispecies as a whole [149]. During viral replication, memory genomes increased their fitness to a similar extent to the entire population to which they belonged [150]. This implies that memory genomes are part of a replicative ensemble, and that their behavior is in agreement with the Red Queen hypothesis [151]. Memory levels are fitness dependent, in that the higher the relative fitness of the genomes destined to become memory genomes, the higher the proportion that they will reach as memory genomes [152]. Behavior in agreement with the Red Queen hypothesis was documented first with VSV [153], and also documented during the evolution of other viruses [146]. Memory genomes can persist for hundreds of infectious cycles, and their decrease in frequency followed a strikingly similar pattern in four parallel lineages [154]. These results indicate that viral quasispecies can manifest deterministic behavior, as observed with the outcome of competitions between VSV mutants [155], and in synchronous alteration of infectivity levels in parallel lineages of cells persistently infected with FMDV [156]. Deterministic behavior has been attributed to an averaging of mutational noise signals that accompany the competitive selection among variants, or to constraints on the types of mutations imposed on compact genomes when directed to evolve in a fixed environment [155,156]. The deterministic behavior during competitions between a wild-type and neutral mutant of VSV [155] illustrated experimentally an additional point on the relevance of the position occupied by a virus in sequence space. In prolonged competitions between the wild-type and mutant VSV, the wild-type, very slowly, always gained dominance over the mutant [155]. However, the introduction of a number of unrelated environmental perturbations (e.g., addition of defective-interfering particles, presence of a mutagenic agent and increase of temperature) led, in all cases, to an accelerated dominance of the wild-type virus [157]. Thus, the neutrality of

the VSV mutant was contingent upon a standard replicative environment. Comparison of the entire consensus genomic sequences showed that the mutant VSV differed from the wild-type in six mutations. The results suggest that the six mutations behaved as neutral mutations under standard replication conditions, but were either deleterious or diminished tolerance to additional mutations when the environment was perturbed. Contingent neutrality when a virus occupies a portion of sequence space in which either beneficial mutations are restricted or deleterious mutations are more frequent (loss of robustness to the effects of mutations) agrees with the concept of advantage of the flattest documented *in silico* and experimentally [157–160].

In addition to offering evidence that the entire quasispecies acts as a unit of selection, the experiments on quasispecies memory and the comparison of parallel evolutionary lineages show that either stochastic or deterministic features can dominate virus evolution. Which of the two prevails may be dependent on a number of parameters, such as virus population size, environmental heterogeneity and position of the virus in the fitness landscape [156,161,162].

The presence of quasispecies memory was extended to HIV-1 populations *in vivo* [100,163]. In this case, two types of memory were distinguished: a replicative memory similar to that described for FMDV [149,152,154], and a ‘cellular’ or ‘anatomical’ memory, derived from integration of proviral DNA into cellular DNA, which results in a reservoir of genomes from previous stages of HIV-1 evolution. Clonal analyses of HIV-1 populations from an infected patient provided evidence that the presence of replicative memory, in the form of genomes encoding multidrug-resistance mutations, influenced the evolution of the HIV-1 quasispecies *in vivo* [164]. The detection of memory genomes associated with drug resistance confers genotyping (by either microarray or ultradeep sequencing procedures), a tool for personalized therapies avoiding administration of drugs whose corresponding resistance mutations are present in the mutant spectrum of the virus to be controlled [164–166].

Replicative memory in viral populations has been interpreted as being a consequence of the population dynamics of RNA viruses, in that the viral replication that follows a selection event increases the relative fitness of the selected viral subpopulations [1,149]. Other influences, such as complementation, which is so relevant to interpreting quasispecies behavior, as discussed in other sections of this article, might also

contribute to maintaining memory levels [167]. The main reason for us to favor quasispecies dynamics as the major memory determinant is that the elongated internal oligoadenylate used as one of the memory markers in FMDV affects *cis*-acting functions, which cannot be sustained by complementation [150]. In summary, quasispecies memory reinforces the value of mutant spectra as determinants of viral evolution and virus behavior, and it opens new prospects in diagnostic virology and the planning of antiviral treatments [165].

Mutant spectrum complexity & viral pathogenesis

Dynamic mutant spectra occur during viral replication *in vivo* (see [3,35–37,48,69,168–172] and several chapters in [40,51,173], among many other examples). Two interpretations of their presence and variations may be considered: that they constitute random mutant distributions fueled by high mutation rates and virus turnover, without implication in virus behavior *in vivo*, or that they are generated by mutation, but that their amplitude and composition is influenced by intra-population interactions and are relevant to virus behavior *in vivo*, including viral pathogenesis. In this section, we summarize some examples of the value of mutant spectra in viral diseases that favor the second interpretation.

The adaptation of a swine FMDV to guinea pig proceeded through heterogeneous viral populations until a key mutation and an accompanying secondary mutation became dominant as the virus increased its pathogenicity for the guinea pig [169]. The relevance of the key mutation that initially appeared as a minority subpopulation was confirmed by introducing it in an infectious viral clone. Interestingly, in this case, the mutant FMDV adapted to cause disease in guinea pigs maintained its capacity to infect swine [174]. Thus, in this case, the mutation mediated an expansion of host range.

Genetic variation of variable genomic regions of HCV has been positively associated with progression towards chronic disease in HCV-infected individuals ([175,176], reviewed in [177]). The transition towards chronicity may involve either immune escape or, alternatively, selection of subpopulations from the mutant spectrum, without necessarily implicating immune escape. Quasispecies dynamics may play a more direct role in maintenance of chronic HCV infections than in their establishment, although this is a debatable issue [74,177,178]. HCV quasispecies complexity at the time of initiation of therapy

influenced the response to standard IFN- α , ribavirin therapy. Low mutant spectrum complexity was associated with sustained viral clearance, while elevated complexity was associated with treatment failure [177,179–181]. Elimination of the infection upon treatment was preceded by a reduction of quasispecies complexity, while treatment failure was frequently linked to changes in the composition of the mutant spectrum [74,176,177,182]. Thus, HCV mutant spectrum composition and dynamics, which should soon become amenable to detailed analysis by ultra-deep sequencing, will offer new prospects for diagnosis and personalized therapy planning.

The fidelity of viral polymerases can be modified by amino acid substitutions in the enzymes, thus providing valuable genetic tools to probe the effects of viral population complexity on virus–host interactions. Studies with a high-fidelity poliovirus mutant – with amino acid substitution G64S in its polymerase – demonstrated that the amplitude of the mutant spectrum was a disease determinant in a model mouse system [48,69]. Low mutant spectrum complexity resulted in limited neuropathology. When the narrow mutant spectrum population was broadened by a mutagenic treatment, the virus regained its capacity to produce disease. It would be generally expected that a broader mutant repertoire may include higher proportions of individuals, which either display a required phenotypic trait or participate in complementation to ensure sufficient fitness for viral replication in the new environments encountered in an infected host. This concept is supported by the observations made with HCV and poliovirus, but the alternative possibility that a narrow mutant spectrum (as a result of a population bottleneck or other factors) might be enriched with virulent subpopulations cannot be excluded. The role of mutant spectrum complexity in addition to changes in the consensus sequence was assigned to the replicative ability of West Nile virus [183,184]. During HIV-1 infections, perhaps due to the multiple interconnected parameters that determine viral evolution, different correlations have been observed. In patients with a moderate-to-slow rate of disease progression, a three-step evolution of divergence from the initial HIV-1 and population diversity, which ended in a stabilization or decline of diversity, was observed [185]. The evolution of diversity was associated with changes in viral co-receptor specificity and in the composition of T-cell subsets. In another study, a low rate of disease progression was associated with a high degree of diversity of the V1, V2,

V3 region of the envelope glycoprotein [186]. In yet another study, HIV-1 from infected patients gained fitness at a rate that correlated with viral load and envelope glycoprotein diversity [187].

Viral quasispecies & viral disease control

One of the most important practical consequences of the viral quasispecies structure is its affect on antiviral therapies. The development of quasispecies in the RNA virus scientific community has had a tremendous impact on the medical treatment of this group of pathogens. Knowledge of HIV-1 variability, and experience of more than a decade with antiretroviral therapy, have taught us that to counteract the large population size, high replication rate and error rates of RNA viruses, antiviral therapies have to involve coadministration of multiple antivirals targeting different viral proteins, or targeting only one viral protein but through different mechanisms of action. This strategy could reduce the emergence of single-resistant viruses, as exemplified with the multiple anti-HIV drug-combination approach, known as HAART [188]. Based on the clinical success of HAART, a similar strategy to counteract viral escape is now being used to treat infections by other RNA viruses, such as HCV and HBV [189]. What currently seems obvious for the treatment of viral infections, was the exception two decades ago, when monotherapy or sequential administration of antiviral compounds were widespread clinical procedures. Therefore, denying the existence of viral quasispecies may have a negative impact on the design of new antiviral strategies. One of the most challenging aims in virology is the development of an effective HIV vaccine to halt the world AIDS pandemic. Interestingly, HIV vaccine development has not kept pace with breakthroughs in basic science. Moreover, HIV quasispecies evolution still remains poorly understood. Without any doubt, a better knowledge of virus quasispecies may help in the development of a more effective HIV vaccine. Of the many HIV/SIV vaccines tested so far, live-attenuated SIV has provided the best protection from pathogenic strains of SIV [190]. Recently, it has been hypothesized that attenuated SIV still retains sufficient replicative capacity to evolve *in vivo* and diversify in response to immune pressure [191]. As a result, the immune system will also evolve and diversify in response to virus evolution. After challenge with a pathogenic but closely related strain of wild-type SIV, the incoming virus will find an immune response selected, not only for its ability to contain the

original virus, but also the preferred immune escape variants of that virus. These conditions resemble those of a drug-combination approach, where multidrug attack is harder to escape than single-drug therapy. This is in contrast with vaccine strategies based on repeated immunization with invariant antigens. However, the capability of live-attenuated SIVs to evolve and diversify makes them too dangerous for use in humans. If this hypothesis is correct, the challenge will be to mimic immune responses induced by attenuated SIV without the use of live retroviruses [191]. What is interesting in this HIV vaccine approach is that virus quasispecies complexity is being taking into account. The need to evoke a broad immune response for vaccine efficacy is broadly recognized [12,42,192,193].

Mutant spectra constitute a repository of phenotypically altered variants that can jeopardize viral disease control by several mechanisms. Pre-existence of mutants resistant to antiviral agents in the absence of antiviral therapy was first described for HIV-1 [194,195], and then extended to other pathogenic viruses. Partially immunized hosts (either naturally or through inefficient vaccines) can promote the selection of antigenic variants. Owing to the frequent overlap between antigenic sites and receptor recognition sites on the virus surface, antigenic variants may contribute viral subpopulations with altered cellular tropism [196–198]. Circulating live-attenuated vaccine viruses may recombine with wild-type viruses, thereby promoting the emergence of new viral forms, an event extremely relevant to viral disease emergence and re-emergence [199,200] (reviews in [51,201,202]).

New tools reveal multiple minority components with biased base composition in mutant spectra

In addition to the standard cloning-sequence analyses and ultradeep sequencing techniques [107–118], procedures for the selective amplification of A,U- or G,C-rich viral RNA genomes have been developed to probe the composition of viral quasispecies [92,203,204]. Its application to several viruses has revealed multiple minority components that contain extreme base compositions, which remain undetected in the standard clonal analyses [203]. A series of A,U-enriched sequences that represent a broad mutant spectrum were characterized in one patient with postvaccinal acute flaccid poliovirus-associated paralysis [203]. This example suggests the possibility of rescuing viral genome derivatives with highly biased base compositions that may be

associated with atypical manifestations of viral infections [7]. The presence of deep mutant spectra with minority components with biased base composition supports the notion that quasispecies constitute reservoirs of profoundly altered virus, and that swarms of defective genomes (either replication competent or otherwise) are present in viral populations. Their frequency may increase as a result of mutagenesis, but they are already abundantly present during replication under standard mutation rates.

Future perspective

Mutant spectra of viral populations (both for RNA and for DNA viruses) will likely be increasingly recognized as an important parameter, whose characterization is needed to understand the biological behavior of viruses (e.g., adaptability, disease potential and host range). In particular, the amplitude and composition of mutant spectra are expected to be revealed as relevant fitness determinants. Careful analyses of mutant spectra may provide information on the prospects that a virus population has of either continued survival or of approaching extinction. The application of ultradeep nucleotide sequencing, and the selective amplification of minority components present in viral populations, should be of great help in revealing the detailed mutant composition of viruses as they replicate and cause disease in their hosts. Analysis of sequential samples should document shifts in population composition that accompany the virus response to internal and external selective constraints (e.g., host antibody and cytotoxic T-cell responses and presence of antiviral agents). Detailed knowledge of the mutant composition of viral populations should provide a useful tool for understanding virus disease emergence, since it will permit identification of potential cell tropism and host range mutants present in mutant spectra. Moreover, for some RNA viruses (e.g., HIV), a better understanding of the viral quasispecies diversity will be necessary to obtain a vaccine that works.

The molecular basis of template copying fidelity by viral RNA-dependent RNA and DNA polymerases should be increasingly understood through the combination of biochemical and structural studies. This will permit altering the copying fidelity of viruses to evaluate the biological consequences of variations in mutation rate with several virus–host systems, following the pioneer studies with poliovirus. This should permit the design of small molecules (nucleoside analogues and other) capable of lowering the fidelity of viral polymerases in a specific manner.

This will greatly expand the possibilities of antiviral interventions using antiviral inhibitors and mutagenic agents either as sequential or combination therapies.

Conclusion, connections, overview & prospects

RNA virus evolution and pathogenesis are decisively influenced by their populations consisting of complex and dynamic mutant spectra, and it would be an incredible step backwards for virology to ignore this fact. Quasispecies were born as a theoretical model for the origin of life [10,11,41], and quasispecies behavior has been shown to apply to ribozyme molecules evolving *in vitro*, and might have contributed to prebiotic RNA molecules eluding extinction [205]. The concept is relevant to highly heterogeneous chemical and biological entities, subjected to the Darwinian principles of competition and selection, including highly variable cell collectivities, such as microbial populations and cancer cells [206–209]. Recently, Darwinian behavior has been documented in prions, based on heterogeneity and selection of conformational states that are determinants of host range, in a process that is reminiscent of quasispecies dynamics [49]. The ample application of the quasispecies concept, and its relevance to several biological cellular and subcellular systems, have been emphasized by Mas and colleagues [15].

Awareness of the quasispecies population structure of viruses is positively guiding the interpretation of clinical data on viral disease and the design of preventive and treatment strategies, as documented in the present article. Deciphering the interactions among components of a mutant spectrum in infected cells will be decisive in interpreting the behavior of viruses at the population level. This will require defining the interaction between wild-type and mutant proteins to form multimeric functional or nonfunctional assemblies, as well as virus–host interactions at the RNA and protein levels. The outcome of such effects is not the result of ‘mystical’ or ‘mysterious’ properties of viral quasispecies (certainly not as mysterious as the unification of electricity, magnetism and light in physics). Quasispecies behavior depends on well-founded concepts in genetics, such as complementation and the action of dominant-negative or interfering mutants, classically tested with defined mutants or defective interfering particles [210]. The same concepts are now applied to the understanding of intramutant spectrum interactions. The molecular interpretation of how such interactions occur in complex quasispecies swarms is a remarkable task still to be

performed. The most advanced molecular biology techniques will be necessary to approach such a formidable challenge. The understanding of quasispecies will move from genetics to biochemistry, the discipline from which the first experimental evidence of quasispecies arose. Such an understanding is critical to resolve issues of virology that relate to the adaptive capacity of complex mutant swarms: persistence, resistance to antiviral agents, accelerated evolution due to vaccination, the emergence and reemergence of viral diseases, possible variant viruses as disease agents yet to be identified and others [6,7,51,73,211–214]. We should not be distracted by semantic or quasisemantic issues. Whether they are termed quasispecies or not, we are dealing with a very unique population structure that has defied antiviral interventions, despite huge economic investments, and that, presently, is helping to interpret virus behavior and to design new means to control viral disease. As Shakespeare said:

“A rose by any other name would smell as sweet.”

Financial & competing interests disclosure

Work at CBMSO, Irsi and Irsi Caixa is supported by grant 360766/09 from Fundación para la Investigación y Prevención del SIDA en España (FIPSE). Work at CBMSO is supported by grants BFU2008–02816/BMC from Ministerio de Ciencia e Innovación and Fundación R. Areces. CIBERehd (Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas) is funded in Instituto de Salud Carlos III. Work at CNM is supported by grants SAF2007–61036, FIPSE 36558/06, 36641/07, 36779/08, and by Red Temática Cooperativa de Investigación en SIDA (Red de grupos 173) of Fondo de Invesigaciones Sanitarias de la Seguridad Social (FISs). Work at Irsi Caixa is supported by grants FIPSE 360766/09 and FIS P107/0098. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

Executive summary

- RNA viruses, and probably also many DNA viruses, replicate as complex mutant distributions, termed ‘viral quasispecies’, both in cell culture and *in vivo*. Many important biological properties of viruses can be missed or misunderstood if the presence of mutant spectra is ignored and only consensus sequences are taken into consideration.
- The presence of mutant clouds is supported by many biochemical studies performed with RNA viruses and their corresponding RNA-dependent RNA and DNA polymerases. For example, in the case of HIV-1, recombination coupled with mutation profoundly influences the evolution of this pathogen during its replication in the human population.
- Mutant spectra of viral populations differ from the classical polymorphic sites described in population genetics. A mutant spectrum has a vastly superior capacity to explore sequence space (and, hence, adaptation) than populations composed of a limited number of alleles.
- The behavior of mutant spectra is influenced by complementations and interferences that are exerted among its individual components. Within-mutant spectrum interactions can give rise to new viral phenotypes. An excess of negative interfering interactions is one of the mechanisms that contribute to virus extinction by lethal mutagenesis.
- Several features of viral quasispecies (e.g., the presence of memory genomes and the striking occasional deterministic behavior of viruses) support the concept of a mutant ensemble being the true target of selection in viruses.
- The mutant spectrum complexity can be an important determinant of viral pathogenesis, as documented by pioneer studies with poliovirus and hepatitis C virus. The amplitude and composition of a mutant spectrum, together with multiple host factors, can influence disease potential and the response to antiviral treatment.
- The understanding of viruses as mutant spectra may contribute to the design of new protocols for antiviral intervention using classic inhibitors and new generations of mutagenic agents, as well as the development of multiepitopic vaccines that take into consideration the adaptive capacity of mutant spectra.
- The new ultradeep nucleotide sequencing and selective amplification techniques offer a means to penetrate into the composition of mutant spectra for more accurate assessment of the potential behavior of viral populations, particularly with regard to treatment planning.

Bibliography

Papers of special note have been highlighted as:

■ of interest

■ of considerable interest

1. Batschelet E, Domingo E, Weissmann C: The proportion of revertant and mutant phage in a growing population, as a function of mutation and growth rate. *Gene* 1(1), 27–32 (1976).

■ First calculation of a mutation rate for an RNA virus.

2. Domingo E, Flavell RA, Weissmann C: *In vitro* site-directed mutagenesis: generation and properties of an infectious extracistronic mutant of bacteriophage Q β . *Gene* 1(1), 3–25 (1976).
3. Domingo E, Sabo D, Taniguchi T, Weissmann C: Nucleotide sequence

heterogeneity of an RNA phage population. *Cell* 13(4), 735–744 (1978).

■ First evidence of quasispecies dynamics in an RNA virus.

4. Domingo E, Holland JJ: High error rates, population equilibrium, and evolution of RNA replication systems. In: *RNA Genetics*. Domingo E, Holland JJ, Ahlquist P (Eds). CRC Press, Boca Raton, FL, USA 3–36 (1988).

5. Domingo E: *Quasispecies: Concept and Implications for Virology (Current Topics in Microbiology and Immunology)*. Springer, Berlin, Germany, 299 (2006).
6. Holland J: Transitions in understanding of RNA viruses: an historical perspective. *Current Topics Microbiol. Immunol.* 299, 371–401 (2006).
7. Holland JJ, Spindler K, Horodyski F, Grabau E, Nichol S, Vandepol S: Rapid evolution of RNA genomes. *Science* 215(4540), 1577–1585 (1982).
- **First account of biological implications of rapid RNA virus evolution.**
8. Novella IS: Contributions of vesicular stomatitis virus to the understanding of RNA virus evolution. *Curr. Opin. Microbiol.* 6(4), 399–405 (2003).
9. Fishman SL, Branch AD: The quasispecies nature and biological implications of the hepatitis C virus. *Infect. Genet. Evol.* 9(6), 1158–1167 (2009).
10. Eigen M: Self-organization of matter and the evolution of biological macromolecules. *Naturwissenschaften* 58, 465–523 (1971).
- **First theoretical treatment of error-prone replication and the origin of genetic information.**
11. Eigen M, Schuster P: *The Hypercycle. A Principle of Natural Self-organization*. Springer, Berlin, Germany (1979).
12. Domingo E, Biebricher C, Eigen M, Holland JJ: *Quasispecies and RNA Virus Evolution: Principles and Consequences*. Landes Bioscience, TX, USA (2001).
13. Domingo E, Holland JJ, Biebricher C, Eigen M: Quasispecies: the concept and the word. In: *Molecular Evolution of the Viruses*. Gibbs A, Calisher C, García-Arenal F (Eds). Cambridge University Press, Cambridge, UK 171–180 (1995).
14. Domingo E, Wain-Hobson S: The 30th anniversary of quasispecies. Meeting on 'Quasispecies: Past, Present and Future'. *EMBO Rep.* 10(5), 444–448 (2009).
15. Mas A, Lopez-Galindez C, Cacho I, Gomez J, Martinez MA: Unfinished stories on viral quasispecies and darwinian views of evolution. *J. Mol. Biol.* 397(4), 865–877 (2010).
- **Review of the broad implications of quasispecies theory for biological systems.**
16. Eigen M: The origin of genetic information: viruses as models. *Gene* 135(1–2), 37–47 (1993).
17. Eigen M: On the nature of virus quasispecies. *Trends. Microbiol.* 4(6), 216–218 (1996).
18. Bull JJ, Meyers LA, Lachmann M: Quasispecies made simple. *PLoS Comput. Biol.* 1(6), E61 (2005).
19. Domingo E: Quasispecies theory in virology. *J. Virol.* 76, 463–465 (2002).
20. Holmes EC, Moya A: Is the quasispecies concept relevant to RNA viruses? *J. Virol.* 76(1), 460–465 (2002).
21. Holmes EC: The RNA virus quasispecies: fact or fiction? *J. Mol. Biol.* 400(3), 271–273 (2010).
22. Eigen M: Natural selection: a phase transition? *Biophys. Chem.* 85, 101–123 (2000).
23. Saakian DB, Hu CK: Exact solution of the Eigen model with general fitness functions and degradation rates. *Proc. Natl Acad. Sci. USA* 103(13), 4935–4939 (2006).
24. Wilke CO, Ronnewinkel C, Martinetz T: Dynamic fitness landscapes in molecular evolution. *Physics Rep.* 349, 395–446 (2001).
25. Wright S: The roles of mutation, inbreeding, crossbreeding, and selection in evolution. *Proc. VI Internatl Congress Genet.* 1, 356–366 (1932).
26. Elena SF, Sanjuan R: Adaptive value of high mutation rates of RNA viruses: separating causes from consequences. *J. Virol.* 79(18), 11555–11558 (2005).
27. Page KM, Nowak MA: Unifying evolutionary dynamics. *J. Theor. Biol.* 219(1), 93–98 (2002).
28. Boerlijst MC, Boenhoefer S, Nowak MA: Viral quasispecies and recombination. *Proc. R. Soc. Lond. B* 263, 1577–1584 (1996).
29. Nowak MA: What is a quasispecies? *Trends in Ecol. Evol.* 4, 118–121 (1992).
30. Nowak MA: *Evolutionary Dynamics*. The Belknap Press of Harvard University Press, MA, USA (2006).
31. Casado C, Pernas M, Alvaro T *et al.*: Coinfection and superinfection in patients with long-term, nonprogressive HIV-1 disease. *J. Infect. Dis.* 196(6), 895–899 (2007).
32. Gottlieb GS, Nickle DC, Jensen MA *et al.*: Dual HIV-1 infection associated with rapid disease progression. *Lancet* 363(9409), 619–622 (2004).
33. Chohan B, Lavreys L, Rainwater SM, Overbaugh J: Evidence for frequent reinfection with human immunodeficiency virus type 1 of a different subtype. *J. Virol.* 79(16), 10701–10708 (2005).
34. Domingo E: Virus evolution. In: *Fields Virology*. Knipe DM, Howley PM (Eds). Wolters Kluwer, Lippincott Williams & Wilkins, PA, USA 389–421 (2007).
35. Gebauer F, De La Torre JC, Gomes I *et al.*: Rapid selection of genetic and antigenic variants of foot-and-mouth disease virus during persistence in cattle. *J. Virol.* 62(6), 2041–2049 (1988).
36. Temin HM: Is HIV unique or merely different? *J. AIDS* 2(1), 1–9 (1989).
37. Schneider WI, Roossinck MJ: Genetic diversity in RNA virus quasispecies is controlled by host–virus interactions. *J. Virol.* 75(14), 6566–6571 (2001).
38. Clements JE, Gdovin SL, Montelaro RC, Narayan O: Antigenic variation in lentiviral diseases. *Annu. Rev. Immunol.* 6, 139–159 (1988).
39. Salinovich O, Payne SL, Montelaro RC, Hussain KA, Issel CJ, Schnorr KL: Rapid emergence of novel antigenic and genetic variants of equine infectious anemia virus during persistent infection. *J. Virol.* 57(1), 71–80 (1986).
40. Domingo E, Brun A, Núñez JI, Cristina J, Briones C, Escarmis C: Genomics of viruses. In: *Pathogenomics: Genome Analysis of Pathogenic Microbes*. Hacker J, Dobrindt U (Eds). Wiley-VCH Verlag GmbH & Co KGaA, Weinheim, Germany 369–388 (2006).
41. Eigen M, Biebricher CK: Sequence space and quasispecies distribution. In: *RNA Genetics*. Domingo E, Ahlquist P, Holland JJ (Eds). CRC Press, FL, USA 211–245 (1988).
42. Domingo E: RNA virus evolution and the control of viral disease. *Prog. Drug Res.* 33, 93–133 (1989).
43. Figlerowicz M, Alejska M, Kurzynska-Kokorniak A, Figlerowicz M: Genetic variability: the key problem in the prevention and therapy of RNA-based virus infections. *Medicinal Res. Reviews* 23(4), 488–518 (2003).
44. Novella IS, Domingo E, Holland JJ: Rapid viral quasispecies evolution: implications for vaccine and drug strategies. *Molecular Med. Today* 1, 248–253 (1995).
45. Ojonegros S, Beerenwinkel N, Antal T, Nowak MA, Escarmis C, Domingo E: Competition–colonization dynamics in an RNA virus. *Proc. Natl Acad. Sci. USA* 107(5), 2108–2112 (2010).
- **Demonstration that a single viral genome can diversify into two distinct subpopulations, whose interaction results in modulation of virulence.**
46. Ojonegros S, Beerenwinkel N, Domingo E: Competition–colonization dynamics: an ecology approach to quasispecies dynamics and virulence evolution in RNA viruses. *Communicative Integrative Biol.* 107(5), 2108–2112 (2010).
47. Stich M, Briones C, Manrubia SC: Collective properties of evolving molecular quasispecies. *BMC Evol. Biol.* 7, 110 (2007).

48. Vignuzzi M, Stone JK, Arnold JJ, Cameron CE, Andino R: Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population. *Nature* 439, 344–348 (2006).
- **Documents the relevance of mutant spectrum complexity in virus adaptability and pathogenesis. Shows quasispecies as a unit of selection *in vivo*.**
49. Li J, Browning S, Mahal SP, Oelschlegel AM, Weissmann C: Darwinian evolution of prions in cell culture. *Science* 327(5967), 869–872 (2010).
- **Darwinian behavior of a nonreplicative population of molecules.**
50. Paul AV: Possible unifying mechanism of picornavirus genome replication. In: *Molecular Biology of Picornaviruses*. Semler BL, Wimmer E (Eds). ASM Press, Washington DC, USA 227–246 (2002).
51. Ehrenfeld E, Domingo E, Ross RP (Eds): *The Picornaviruses*. ASM Press, Washington DC, USA (2010).
52. Lyle JM, Bullitt E, Bienz K, Kirkegaard K: Visualization and functional analysis of RNA-dependent RNA polymerase lattices. *Science* 296(5576), 2218–2222 (2002).
53. Parvin JD, Moscona A, Pan WT, Leider JM, Palese P: Measurement of the mutation rates of animal viruses: influenza a virus and poliovirus type 1. *J. Virol.* 59(2), 377–383 (1986).
54. Emini EA, Jameson BA, Lewis AJ, Larsen GR, Wimmer E: Poliovirus neutralization epitopes: analysis and localization with neutralizing monoclonal antibodies. *J. Virol.* 43(3), 997–1005 (1982).
55. Minor PD, Schild GC, Bootman J *et al.*: Location and primary structure of a major antigenic site for poliovirus neutralization. *Nature* 301(5902), 674–679 (1983).
56. Minor PD, Ferguson M, Evans DM, Almond JW, Icenogle JP: Antigenic structure of polioviruses of serotypes 1, 2 and 3. *J. Gen. Virol.* 67, 1283–1291 (1986).
57. Sedivy JM, Capone JP, Rajbhandary UL, Sharp PA: An inducible mammalian amber suppressor: propagation of a poliovirus mutant. *Cell* 50(3), 379–389 (1987).
58. De La Torre JC, Wimmer E, Holland JJ: Very high frequency of reversion to guanidine resistance in clonal pools of guanidine-dependent type 1 poliovirus. *J. Virol.* 64(2), 664–671 (1990).
59. Ward CD, Flanagan JB: Determination of the poliovirus RNA polymerase error frequency at eight sites in the viral genome. *J. Virol.* 66(6), 3784–3793 (1992).
60. Steinhauer DA, Holland JJ: Direct method for quantitation of extreme polymerase error frequencies at selected single base sites in viral RNA. *J. Virol.* 57(1), 219–228 (1986).
61. Domingo E, Holland JJ: Mutation rates and rapid evolution of RNA viruses. In: *Evolutionary Biology of Viruses*. Morse SS (Ed.). Raven Press, NY, USA 161–184 (1994).
62. Ng KK, Arnold JJ, Cameron CE: Structure–function relationships among RNA-dependent RNA polymerases. *Curr. Topics Microbiol. Immunol.* 320, 137–156 (2008).
63. Arnold JJ, Vignuzzi M, Stone JK, Andino R, Cameron CE: Remote site control of an active site fidelity checkpoint in a viral rna-dependent RNA polymerase. *J. Biol. Chem.* 280(27), 25706–25716 (2005).
64. Ferrer-Orta C, Arias A, Perez-Luque R, Escarmis C, Domingo E, Verdaguier N: Sequential structures provide insights into the fidelity of RNA replication. *Proc. Natl Acad. Sci. USA* 104(22), 9463–9468 (2007).
- **Structural study of the interaction of a virus polymerase with nucleotides and nucleotide analogues.**
65. Menéndez-Arias L: Molecular basis of fidelity of DNA synthesis and nucleotide specificity of retroviral reverse transcriptases. *Prog. Nucl. Acid Res. Mol. Biol.* 71, 91–147 (2002).
66. Castro C, Arnold JJ, Cameron CE: Incorporation fidelity of the viral RNA-dependent RNA polymerase: A kinetic, thermodynamic and structural perspective. *Virus Res.* 107(2), 141–149 (2005).
67. Arias A, Arnold JJ, Sierra M, Smidansky ED, Domingo E, Cameron CE: Determinants of RNA-dependent RNA polymerase (in)fidelity revealed by kinetic analysis of the polymerase encoded by a foot-and-mouth disease virus mutant with reduced sensitivity to ribavirin. *J. Virol.* 82(24), 12346–12355 (2008).
68. Jiang C, Komazin-Meredith G, Tian W, Coen DM, Hwang CB: Mutations that increase DNA binding by the processivity factor of herpes simplex virus affect virus production and DNA replication fidelity. *J. Virol.* 83(15), 7573–7580 (2009).
69. Pfeiffer JK, Kirkegaard K: Increased fidelity reduces poliovirus fitness under selective pressure in mice. *PLoS Pathog.* 1, 102–110 (2005).
- **Documents the relevance of mutant spectrum complexity in virus adaptability and pathogenesis. Shows quasispecies as a unit of selection *in vivo*.**
70. Earl DJ, Deem MW: Evolvability is a selectable trait. *Proc. Natl Acad. Sci. USA* 101(32), 11531–11536 (2004).
71. Sanjuan R: Mutational fitness effects in RNA and single-stranded DNA viruses: common patterns revealed by site-directed mutagenesis studies. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 365(1548), 1975–1982 (2010).
72. Mims C, Nash A, Stephen J: *Mims' Pathogenesis of Infectious Disease*. Academic Press, CA, USA (2001).
73. Domingo E, Escarmis C, Menéndez-Arias L *et al.*: Viral quasispecies: dynamics, interactions and pathogenesis. In: *Origin and Evolution of Viruses*. Domingo E, Parrish C, Holland JJ (Eds). Elsevier, Oxford, UK 87–118 (2008).
74. Domingo E, Gomez J: Quasispecies and its impact on viral hepatitis. *Virus Res.* 127, 131–150 (2007).
75. Domingo E, Grande-Pérez A, Martin V: Future prospects for the treatment of rapidly evolving viral pathogens: insights from evolutionary biology. *Expert Opin. Biol. Ther.* 8(10), 1455–1460 (2008).
76. Bernad A, Blanco L, Lazaro JM, Martin G, Salas M: A conserved 3'–5' exonuclease active site in prokaryotic and eukaryotic DNA polymerases. *Cell* 59(1), 219–228 (1989).
77. Ferrer-Orta C, Arias A, Escarmis C, Verdaguier N: A comparison of viral RNA-dependent RNA polymerases. *Curr. Opin. Struct. Biol.* 16(1), 27–34 (2006).
78. Steinhauer DA, Domingo E, Holland JJ: Lack of evidence for proofreading mechanisms associated with an RNA virus polymerase. *Gene* 122(2), 281–288 (1992).
79. Eckerle LD, Lu X, Sperry SM, Choi L, Denison MR: High fidelity of murine hepatitis virus replication is decreased in nsp14 exonuclease mutants. *J. Virol.* 81(22), 12135–12144 (2007).
80. Eckerle LD, Becker MM, Halpin RA *et al.*: Infidelity of SARS-Cov nsp14-exonuclease mutant virus replication is revealed by complete genome sequencing. *PLoS Pathog.* 6(5), E1000896 (2010).
81. Menendez-Arias L: Targeting HIV: antiretroviral therapy and development of drug resistance. *Trends Pharmacol. Sci.* 23(8), 381–388 (2002).
82. Friedberg EC, Walker GC, Siede W, Wood RD, Schultz RA, Ellenberger T: *DNA Repair and Mutagenesis*. American Society for Microbiology, Washington, DC, USA (2006).
83. Pita JS, De Miranda JR, Schneider WL, Roossinck MJ: Environment determines fidelity for an RNA virus replicase. *J. Virol.* 81(17), 9072–9077 (2007).

84. Mansky LM, Temin HM: Lower *in vivo* mutation rate of human immunodeficiency virus type 1 than that predicted from the fidelity of purified reverse transcriptase. *J. Virol.* 69(8), 5087–5094 (1995).
85. Bertram JG, Oertel K, Petruska J, Goodman MF: DNA polymerase fidelity: comparing direct competition of right and wrong dNTP substrates with steady state and pre-steady state kinetics. *Biochemistry* 49(1), 20–28 (2010).
86. Jung A, Maier R, Vartanian JP *et al.*: Multiply infected spleen cells in HIV patients. *Nature* 418(6894), 144 (2002).
87. Jetzt AE, Yu H, Klarmann GJ, Ron Y, Preston BD, Dougherty JP: High rate of recombination throughout the human immunodeficiency virus type 1 genome. *J. Virol.* 74(3), 1234–1240 (2000).
88. Meyerhans A, Jung A, Maier R, Vartanian JP, Bocharov G, Wain-Hobson S: The non-clonal and transitory nature of HIV *in vivo*. *Swiss Med. Wkly* 133(33–34), 451–454 (2003).
89. Quan Y, Liang C, Brenner BG, Wainberg MA: Multidrug-resistant variants of HIV type 1 (HIV-1) can exist in cells as defective quasiespecies and be rescued by superinfection with other defective HIV-1 variants. *J. Infect. Dis.* 200(9), 1479–1483 (2009).
90. Li Y, Kappes JC, Conway JA, Price RW, Shaw GM, Hahn BH: Molecular characterization of human immunodeficiency virus type 1 cloned directly from uncultured human brain tissue: Identification of replication-competent and -defective viral genomes. *J. Virol.* 65(8), 3973–3985 (1991).
91. Sadler HA, Stenglein MD, Harris RS, Mansky LM: APOBEC3G contributes to HIV-1 variation through sublethal mutagenesis. *J. Virol.* 84(14), 7396–7404 (2010).
92. Suspene R, Guetard D, Henry M, Sommer P, Wain-Hobson S, Vartanian JP: Extensive editing of both hepatitis B virus DNA strands by APOBEC3 cytidine deaminases *in vitro* and *in vivo*. *Proc. Natl Acad. Sci. USA* 102(23), 8321–8326 (2005).
93. Vartanian JP, Henry M, Marchio A *et al.*: Massive APOBEC3 editing of hepatitis B viral DNA in cirrhosis. *PLoS Pathog.* 6(5), e1000928 (2010).
94. Martínez MA, Vartanian JP, Wain-Hobson S: Hypermutagenesis of RNA using human immunodeficiency virus type 1 reverse transcriptase and biased dNTP concentrations. *Proc. Natl Acad. Sci. USA* 91(25), 11787–11791. (1994).
95. Cheynier R, Henrichwark S, Hadida F *et al.*: HIV and T cell expansion in splenic white pulps is accompanied by infiltration of HIV-specific cytotoxic T lymphocytes. *Cell* 78(3), 373–387 (1994).
96. Korber BT, Kunstman KJ, Patterson BK *et al.*: Genetic differences between blood- and brain-derived viral sequences from human immunodeficiency virus type 1-infected patients: evidence of conserved elements in the V3 region of the envelope protein of brain-derived sequences. *J. Virol.* 68(11), 7467–7481 (1994).
97. Wong JK, Ignacio CC, Torriani F, Havlir D, Fitch NJ, Richman DD: *In vivo* compartmentalization of human immunodeficiency virus: evidence from the examination of pol sequences from autopsy tissues. *J. Virol.* 71(3), 2059–2071 (1997).
98. Sala M, Zambruno G, Vartanian JP, Marconi A, Bertazzoni U, Wain-Hobson S: Spatial discontinuities in human immunodeficiency virus type 1 quasiespecies derived from epidermal Langerhans cells of a patient with AIDS and evidence for double infection. *J. Virol.* 68(8), 5280–5283 (1994).
99. Dobzhansky T, Ayala FJ, Stebbins GL, Valentine JW: *Evolution*. WH Freeman and company, CA, USA (1977).
100. Quan Y, Brenner BG, Dascal A, Wainberg MA: Highly diversified multiply drug-resistant HIV-1 quasiespecies in PBMCs: a case report. *Retrovirology* 5, 43 (2008).
101. Spiess EB: *Genes in Populations*. John Wiley & Sons, NY, USA (1977).
102. Hunter RL, Markert CL: Histochemical demonstration of enzymes separated by zone electrophoresis in starch gels. *Science* 125(3261), 1294–1295 (1957).
103. Markert CL, Moller F: Multiple forms of enzymes: tissue, ontogenetic, and species specific patterns. *Proc. Natl Acad. Sci. USA* 45(5), 753–763 (1959).
104. Harris H: Enzyme polymorphisms in man. *Proc. R. Soc. Lond. B Biol. Sci.* 164(995), 298–310 (1966).
105. Lewontin RC, Hubby JL: A molecular approach to the study of genic heterozygosity in natural populations. II. Amount of variation and degree of heterozygosity in natural populations of *Drosophila pseudoobscura*. *Genetics* 54(2), 595–609 (1966).
106. Hubby JL, Lewontin RC: A molecular approach to the study of genic heterozygosity in natural populations. I. The number of alleles at different loci in *Drosophila pseudoobscura*. *Genetics* 54(2), 577–594 (1966).
107. Ghedin E, Sengamalay NA, Shumway M *et al.*: Large-scale sequencing of human influenza reveals the dynamic nature of viral genome evolution. *Nature* 437(7062), 1162–1166 (2005).
108. Eriksson N, Pachter L, Mitsuya Y *et al.*: Viral population estimation using pyrosequencing. *PLoS Comput. Biol.* 4(4), E1000074 (2008).
109. Tsibris AM, Korber B, Arnaout R *et al.*: Quantitative deep sequencing reveals dynamic HIV-1 escape and large population shifts during CCR5 antagonist therapy *in vivo*. *PLoS One* 4(5), E5683 (2009).
110. Wang C, Mitsuya Y, Gharizadeh B, Ronaghi M, Shafer RW: Characterization of mutation spectra with ultra-deep pyrosequencing: application to HIV-1 drug resistance. *Genome Res.* 17(8), 1195–1201 (2007).
111. Mitsuya Y, Varghese V, Wang C *et al.*: Minority human immunodeficiency virus type 1 variants in antiretroviral-naïve persons with reverse transcriptase codon 215 revertant mutations. *J. Virol.* 82(21), 10747–10755 (2008).
112. Hoffmann C, Minkah N, Leipzig J *et al.*: DNA bar coding and pyrosequencing to identify rare HIV drug resistance mutations. *Nucleic Acids Res.* 35(13), E91 (2007).
113. Kuroda M, Katano H, Nakajima N *et al.*: Characterization of quasiespecies of pandemic 2009 influenza A virus (A/H1N1/2009) by de novo sequencing using a next-generation DNA sequencer. *PLoS One* 5(4), e10256 (2010).
114. Ramakrishnan MA, Tu ZJ, Singh S *et al.*: The feasibility of using high resolution genome sequencing of influenza A viruses to detect mixed infections and quasiespecies. *PLoS One* 4(9), E7105 (2009).
115. Webster DR, Hekele AG, Lauring AS *et al.*: An enhanced single base extension technique for the analysis of complex viral populations. *PLoS One* 4(10), E7453 (2009).
116. Margeridon-Thermet S, Shulman NS, Ahmed A *et al.*: Ultra-deep pyrosequencing of hepatitis B virus quasiespecies from nucleoside and nucleotide reverse-transcriptase inhibitor (NRTI)-treated patients and NRTI-naïve patients. *J. Infect. Dis.* 199(9), 1275–1285 (2009).
117. Solmone M, Vincenti D, Prosperi MC, Bruselles A, Ippolito G, Capobianchi MR: Use of massively parallel ultradeep pyrosequencing to characterize the genetic diversity of hepatitis B virus in drug-resistant and drug-naïve patients and to detect minor variants in reverse transcriptase and hepatitis B S antigen. *J. Virol.* 83(4), 1718–1726 (2009).

118. Zagordi O, Klein R, Daumer M, Beerenwinkel N: Error correction of next-generation sequencing data and reliable estimation of HIV quasispecies. *Nucleic Acids Res.* doi: 10.1093/nar/gkq655 (2010) (Epub ahead of print).
119. Igarashi M, Ito K, Yoshida R, Tomabechi D, Kida H, Takada A: Predicting the antigenic structure of the pandemic (H1N1) 2009 influenza virus hemagglutinin. *PLoS One* 5(1), E8553 (2009).
120. Shen J, Ma J, Wang Q: Evolutionary trends of A(H1N1) influenza virus hemagglutinin since 1918. *PLoS One* 4(11), E7789 (2009).
121. Fazekas De St Groth S, Hannoun C: Selection of dominant mutants of the influenza A (hong kong) virus by immunological pressure. *C. R. Acad. Sci. Hebd. Seances Acad. Sci. D* 276(12), 1917–1920 (1973).
122. Webster RG, Laver WG: Antigenic variation of influenza viruses. In: *The Influenza Viruses and Influenza*. Kilbourne ED (Ed.). Academic Press, NY, USA 269–314 (1975).
123. Burch CL, Chao L: Evolvability of an RNA virus is determined by its mutational neighbourhood. *Nature* 406(6796), 625–628 (2000).
124. De La Torre JC, Holland JJ: RNA virus quasispecies populations can suppress vastly superior mutant progeny. *J. Virol.* 64(12), 6278–6281 (1990).
- **First demonstration of suppression of a high-fitness virus variant by a mutant spectrum.**
125. Swetina J, Schuster P: Self-replication with errors. A model for polynucleotide replication. *Biophys. Chem.* 16(4), 329–345 (1982).
126. Perales C, Martin V, Ruiz-Jarabo CM, Domingo E: Monitoring sequence space as a test for the target of selection in viruses. *J. Mol. Biol.* 345(3), 451–459 (2005).
127. Eigen M, McCaskill J, Schuster P: Molecular quasi-species. *J. Phys. Chem.* 92, 6881–6891 (1988).
128. González-López C, Arias A, Pariente N, Gómez-Mariano G, Domingo E: Preextinction viral RNA can interfere with infectivity. *J. Virol.* 78(7), 3319–3324 (2004).
- **Evidence of suppression of viral infectivity by a mutagenized viral population.**
129. Domingo E, Horzinek MC: Epilogue. Animal virology: a showcase of evolution. In: *Animal Viruses. Molecular Biology*. Mettenleiter TC, Sobrino F (Eds).
130. Caister Academic Press, Norfolk, UK 523–531 (2008). Crowder S, Kirkegaard K: Trans-dominant inhibition of RNA viral replication can slow growth of drug-resistant viruses. *Nature Genetics* 37(7), 701–709 (2005).
- **Demonstration of internal interactions within mutant spectra.**
131. Grande-Pérez A, Lazaro E, Lowenstein P, Domingo E, Manrubia SC: Suppression of viral infectivity through lethal defection. *Proc. Natl Acad. Sci. USA* 102(12), 4448–4452 (2005).
- **First formulation of the lethal defection model of virus extinction.**
132. Irazzo J, Manrubia SC: Stochastic extinction of viral infectivity through the action of defectors. *Europhys. Lett.* 85, 18001 (2009).
133. Arias A, Agudo R, Ferrer-Orta C *et al.*: Mutant viral polymerase in the transition of virus to error catastrophe identifies a critical site for RNA binding. *J. Mol. Biol.* 353(5), 1021–1032 (2005).
134. Mateo R, Diaz A, Baranowski E, Mateu MG: Complete alanine scanning of intersubunit interfaces in a foot-and-mouth disease virus capsid reveals critical contributions of many side chains to particle stability and viral function. *J. Biol. Chem.* 278(42), 41019–41027 (2003).
135. Perales C, Mateo R, Mateu MG, Domingo E: Insights into RNA virus mutant spectrum and lethal mutagenesis events: replicative interference and complementation by multiple point mutants. *J. Mol. Biol.* 369(4), 985–1000 (2007).
- **Demonstration of complementation and interference by specific viral mutants.**
136. Duarte EA, Novella IS, Ledesma S *et al.*: Subclonal components of consensus fitness in an RNA virus clone. *J. Virol.* 68(7), 4295–4301 (1994).
137. Chumakov KM, Powers LB, Noonan KE, Roninson IB, Levenbook IS: Correlation between amount of virus with altered nucleotide sequence and the monkey test for acceptability of oral poliovirus vaccine. *Proc. Natl Acad. Sci. USA* 88(1), 199–203 (1991).
138. Teng MN, Oldstone MB, De La Torre JC: Suppression of lymphocytic choriomeningitis virus-induced growth hormone deficiency syndrome by disease-negative virus variants. *Virology* 223(1), 113–119 (1996).
139. Borrego B, Novella IS, Giralte E, Andreu D, Domingo E: Distinct repertoire of antigenic variants of foot-and-mouth disease virus in the presence or absence of immune selection. *J. Virol.* 67(10), 6071–6079 (1993).
140. Moreno IM, Malpica JM, Rodríguez-Cerezo E, García-Arenal F: A mutation in tomato aspermy cucumovirus that abolishes cell-to-cell movement is maintained to high levels in the viral RNA population by complementation. *J. Virol.* 71(12), 9157–9162 (1997).
141. Askov J, Buzacott K, Thu HM, Lowry K, Holmes EC: Long-term transmission of defective RNA viruses in humans and aedes mosquitoes. *Science* 311(5758), 236–238 (2006).
142. Borderia AV, Lorenzo-Redondo R, Pernas M *et al.*: Initial fitness recovery of HIV-1 is associated with quasispecies heterogeneity and can occur without modifications in the consensus sequence. *PLoS One* 5(4), e10319 (2010).
- **Evidence that fitness gain of HIV-1 can occur with an invariant consensus sequence.**
143. González-López C, Gómez-Mariano G, Escarmis C, Domingo E: Invariant aphthovirus consensus nucleotide sequence in the transition to error catastrophe. *Infection Genet. Evolution* 5, 366–374 (2005).
144. Sanz-Ramos M, Diaz-San Segundo F, Escarmis C, Domingo E, Sevilla N: Hidden virulence determinants in a viral quasispecies *in vivo*. *J. Virol.* 82(21), 10465–10476 (2008).
145. Yang Z, Nielsen R: Mutation-selection models of codon substitution and their use to estimate selective strengths on codon usage. *Mol. Biol. Evol.* 25(3), 568–579 (2008).
146. Aragones L, Guix S, Ribes E, Bosch A, Pinto RM: Fine-tuning translation kinetics selection as the driving force of codon usage bias in the hepatitis A virus capsid. *PLoS Pathog.* 6(3), E1000797 (2010).
147. Jerne NK: The natural-selection theory of antibody formation. *Proc. Natl Acad. Sci. USA* 41(11), 849–857 (1955).
148. Frank SA: The design of natural and artificial adaptive systems. In: *Adaptation*. Rose MR, Lauder GV (Eds). Academic Press, CA, USA 451–505 (1996).
149. Ruiz-Jarabo CM, Arias A, Baranowski E, Escarmis C, Domingo E: Memory in viral quasispecies. *J. Virol.* 74, 3543–3547 (2000).
- **Demonstration of the presence of memory genomes in viral quasispecies.**
150. Arias A, Ruiz-Jarabo CM, Escarmis C, Domingo E: Fitness increase of memory genomes in a viral quasispecies. *J. Mol. Biol.* 339(2), 405–412 (2004).
151. Van Valen L: A new evolutionary law. *Evol. Theory* 1, 1–30 (1973).
152. Ruiz-Jarabo CM, Arias A, Molina-París C *et al.*: Duration and fitness dependence of quasispecies memory. *J. Mol. Biol.* 315(3), 285–296 (2002).

153. Clarke DK, Duarte EA, Elena SF, Moya A, Domingo E, Holland J: The Red Queen reigns in the kingdom of RNA viruses. *Proc. Natl Acad. Sci. USA* 91(11), 4821–4824 (1994).
154. Ruiz-Jarabo CM, Miller E, Gómez-Mariano G, Domingo E: Synchronous loss of quasispecies memory in parallel viral lineages: a deterministic feature of viral quasispecies. *J. Mol. Biol.* 333, 553–563 (2003).
155. Quer J, Huerta R, Novella IS, Tsimring L, Domingo E, Holland JJ: Reproducible nonlinear population dynamics and critical points during replicative competitions of RNA virus quasispecies. *J. Mol. Biol.* 264(3), 465–471 (1996).
156. Herrera M, Grande-Pérez A, Perales C, Domingo E: Persistence of foot-and-mouth disease virus in cell culture revisited: implications for contingency in evolution. *J. Gen. Virol.* 89 (Pt 1), 232–244 (2008).
157. Quer J, Hershey CL, Domingo E, Holland JJ, Novella IS: Contingent neutrality in competing viral populations. *J. Virol.* 75(16), 7315–7320 (2001).
158. Wilke CO, Wang JL, Ofria C, Lenski RE, Adami C: Evolution of digital organisms at high mutation rates leads to survival of the flattest. *Nature* 412(6844), 331–333 (2001).
159. Codoner FM, Daros JA, Sole RV, Elena SF: The fittest versus the flattest: experimental confirmation of the quasispecies effect with subviral pathogens. *PLoS Pathog.* 2(12), E136 (2006).
160. Sanjuan R, Cuevas JM, Furio V, Holmes EC, Moya A: Selection for robustness in mutagenized RNA viruses. *PLoS Genet.* 3(6), E93 (2007).
161. Lazaro E, Escarmis C, Perez-Mercader J, Manrubia SC, Domingo E: Resistance of virus to extinction on bottleneck passages: study of a decaying and fluctuating pattern of fitness loss. *Proc. Natl Acad. Sci. USA* 100(19), 10830–10835 (2003).
162. Rouzine IM, Rodrigo A, Coffin JM: Transition between stochastic evolution and deterministic evolution in the presence of selection: general theory and application to virology. *Microbiol. Mol. Biol. Rev.* 65(1), 151–185 (2001).
163. Briones C, Domingo E, Molina-París C: Memory in retroviral quasispecies: experimental evidence and theoretical model for human immunodeficiency virus. *J. Mol. Biol.* 331(1), 213–229 (2003).
164. Briones C, De Vicente A, Molina-Paris C, Domingo E: Minority memory genomes can influence the evolution of HIV-1 quasispecies *in vivo*. *Gene* 384, 129–138 (2006).
165. Briones C, Domingo E: Minority report: Hidden memory genomes in HIV-1 quasispecies and possible clinical implications. *AIDS Rev.* 10(2), 93–109 (2008).
166. Domingo E, Ruiz-Jarabo CM, Arias A *et al.*: Detection and biological implications of genetic memory in viral quasispecies. In: *Cardiomyopathies and Heart Failure: Biomolecular, Infectious and Immune Mechanisms*. Matsumori A (Ed.). Kluwer Academic Publishers, London, UK, 259–276 (2003).
167. Wilke CO, Novella IS: Phenotypic mixing and hiding may contribute to memory in viral quasispecies. *BMC Microbiol.* 3(1), 11 (2003).
168. Roossinck MJ: Mutant clouds and bottleneck events in plant virus evolution. In: *Origin and Evolution of Viruses*. Domingo E, Parrish C, Holland JJ (Eds). Elsevier, Oxford, UK 251–258 (2008).
169. Núñez JI, Baranowski E, Molina N *et al.*: A single amino acid substitution in nonstructural protein 3A can mediate adaptation of foot-and-mouth disease virus to the guinea pig. *J. Virol.* 75(8), 3977–3983 (2001).
170. Carrillo C, Plana J, Mascarella R, Bergada J, Sobrino F: Genetic and phenotypic variability during replication of foot-and-mouth disease virus in swine. *Virology* 179(2), 890–892 (1990).
171. Carrillo C, Lu Z, Borca MV, Vagnozzi A, Kutish GF, Rock DL: Genetic and phenotypic variation of foot-and-mouth disease virus during serial passages in a natural host. *J. Virol.* 81(20), 11341–11351 (2007).
172. Shao W, Kearney M, Maldarelli F *et al.*: RT-SHIV subpopulation dynamics in infected macaques during anti-HIV therapy. *Retrovirology* 6, 101 (2009).
173. Domingo E, Parrish C, Holland JJ: *Origin and Evolution of Viruses*. 2nd Edition. Elsevier, Oxford, UK (2008).
174. Núñez JI, Molina N, Baranowski E *et al.*: Guinea pig-adapted foot-and-mouth disease virus with altered receptor recognition can productively infect a natural host. *J. Virol.* 81(16), 8497–8506 (2007).
175. Farci P, Shimoda A, Coiana A *et al.*: The outcome of acute hepatitis C predicted by the evolution of the viral quasispecies. *Science* 288(5464), 339–344 (2000).
176. Farci P, Quinti I, Farci S *et al.*: Evolution of hepatitis C viral quasispecies and hepatic injury in perinatally infected children followed prospectively. *Proc. Natl Acad. Sci. USA* 103(22), 8475–8480 (2006).
177. Pawlotsky JM: Hepatitis C virus population dynamics during infection. *Current Topics Microbiol. Immunol.* 299, 261–284 (2006).
178. Bowen DG, Walker CM: The origin of quasispecies: cause or consequence of chronic hepatitis C viral infection? *J. Hepatol.* 42(3), 408–417 (2005).
179. Pawlotsky JM, Pellerin M, Bouvier M *et al.*: Genetic complexity of the hypervariable region 1 (HVR1) of hepatitis C virus (HCV): influence on the characteristics of the infection and responses to interferon α therapy in patients with chronic hepatitis C. *J. Med. Virol.* 54(4), 256–264 (1998).
180. Pawlotsky JM: Hepatitis C virus resistance to antiviral therapy. *Hepatology* 32(5), 889–896 (2000).
181. Pawlotsky JM: Mechanisms of antiviral treatment efficacy and failure in chronic hepatitis C. *Antiviral Res.* 59(1), 1–11 (2003).
182. Farci P, Strazzera R, Alter HJ *et al.*: Early changes in hepatitis C viral quasispecies during interferon therapy predict the therapeutic outcome. *Proc. Natl Acad. Sci. USA* 99(5), 3081–3086 (2002).
183. Ciota AT, Ngo KA, Lovelace AO *et al.*: Role of the mutant spectrum in adaptation and replication of West Nile virus. *J. Gen. Virol.* 88(Pt 3), 865–874 (2007).
184. Jerzak G, Bernard KA, Kramer LD, Ebel GD: Genetic variation in West Nile virus from naturally infected mosquitoes and birds suggests quasispecies structure and strong purifying selection. *J. Gen. Virol.* 86(Pt 8), 2175–2183 (2005).
185. Shankarappa R, Margolick JB, Gange SJ *et al.*: Consistent viral evolutionary changes associated with the progression of human immunodeficiency virus type 1 infection. *J. Virol.* 73(12), 10489–10502 (1999).
- **Consistent model that relates viral evolution and pathogenesis in HIV-1.**
186. Shioda T, Oka S, Xin X *et al.*: *In vivo* sequence variability of human immunodeficiency virus type 1 envelope gp120: association of V2 extension with slow disease progression. *J. Virol.* 71(7), 4871–4881 (1997).
187. Troyer RM, Collins KR, Abraha A *et al.*: Changes in human immunodeficiency virus type 1 fitness and genetic diversity during disease progression. *J. Virol.* 79(14), 9006–9018 (2005).
- **First correlation between HIV-1 fitness and population heterogeneity.**
188. Ho DD: Time to hit HIV, early and hard. *N. Engl. J. Med.* 333(7), 450–451 (1995).
189. Monto A, Schooley RT, Lai JC *et al.*: Lessons from HIV therapy applied to viral hepatitis therapy: summary of a workshop. *Am. J. Gastroenterol.* 105(5), 989–1004 (2010).

190. Koff WC, Johnson PR, Watkins DI *et al.*: HIV vaccine design: insights from live attenuated SIV vaccines. *Nat. Immunol.* 7(1), 19–23 (2006).
191. Cohen GB: Mechanism of protection of live attenuated simian immunodeficiency virus: coevolution of viral and immune responses. *AIDS* 24(5), 637–648 (2010).
192. Fischer W, Perkins S, Theiler J *et al.*: Polyvalent vaccines for optimal coverage of potential T-cell epitopes in global HIV-1 variants. *Nat. Med.* 13(1), 100–106 (2007).
193. Barouch DH, O'Brien KL, Simmons NL *et al.*: Mosaic HIV-1 vaccines expand the breadth and depth of cellular immune responses in rhesus monkeys. *Nat. Med.* 16(3), 319–323 (2010).
194. Nájera I, Holguín A, Quiñones-Mateu ME *et al.*: Pol gene quasispecies of human immunodeficiency virus: mutations associated with drug resistance in virus from patients undergoing no drug therapy. *J. Virol.* 69(1), 23–31 (1995).
195. Nájera I, Richman DD, Olivares I *et al.*: Natural occurrence of drug resistance mutations in the reverse transcriptase of human immunodeficiency virus type 1 isolates. *AIDS Res. Hum. Retroviruses* 10(11), 1479–1488 (1994).
196. Tami C, Taboga O, Berinstein A *et al.*: Evidence of the coevolution of antigenicity and host cell tropism of foot-and-mouth disease virus *in vivo*. *J. Virol.* 77(2), 1219–1226 (2003).
197. Baranowski E, Ruiz-Jarabo CM, Domingo E: Evolution of cell recognition by viruses. *Science* 292(5519), 1102–1105 (2001).
198. Baranowski E, Ruiz-Jarabo CM, Pariente N, Verdager N, Domingo E: Evolution of cell recognition by viruses: a source of biological novelty with medical implications. *Adv. Virus Res.* 62, 19–111 (2003).
199. Georgescu MM, Balanant J, Macadam A *et al.*: Evolution of the Sabin type 1 poliovirus in humans: characterization of strains isolated from patients with vaccine-associated paralytic poliomyelitis. *J. Virol.* 71(10), 7758–7768 (1997).
200. Chong YL, Padhi A, Hudson PJ, Poss M: The effect of vaccination on the evolution and population dynamics of avian paramyxovirus-1. *PLoS Pathog.* 6(4), E1000872 (2010).
201. Agol VI: Molecular mechanisms of poliovirus variation and evolution. *Current Topics Microbiol. Immunol.* 299, 211–259 (2006).
202. Kew OM, Wright PF, Agol VI *et al.*: Circulating vaccine-derived polioviruses: current state of knowledge. *Bull. World Health Organ.* 82(1), 16–23 (2004).
203. Suspene R, Henry M, Guillot S, Wain-Hobson S, Vartanian JP: Recovery of APOBEC3-edited human immunodeficiency virus G→A hypermutants by differential DNA denaturation PCR. *J. Gen. Virol.* 86(Pt 1), 125–129 (2005).
204. Suspene R, Renard M, Henry M *et al.*: Inverting the natural hydrogen bonding rule to selectively amplify GC-rich ADAR-edited RNAs. *Nucleic Acids Res.* 36(12), E72 (2008).
205. Diaz Arenas C, Lehman N: Quasispecies-like behavior observed in catalytic RNA populations evolving in a test tube. *BMC Evol. Biol.* 10, 80 (2010).
206. Nicolson GL: Tumor cell instability, diversification, and progress on to the metastatic phenotype: from oncogene to oncofetal expression. *Cancer Res.* 47(6), 1473–1487 (1987).
207. Brumer Y, Michor F, Shakhnovich EI: Genetic instability and the quasispecies model. *J. Theor. Biol.* 241(2), 216–222 (2006).
208. Gonzalez-Garcia I, Sole RV, Costa J: Metapopulation dynamics and spatial heterogeneity in cancer. *Proc. Natl Acad. Sci. USA* 99(20), 13085–13089 (2002).
209. Solé RV, Deisboeck TS: An error catastrophe in cancer? *J. Theor. Biol.* 228(1), 47–54 (2004).
210. Condit RC: Principles of virology. In: *Fields Virology*. Knipe DM, Howley PM (Eds). Wolters Kluwer, Lippincott Williams & Wilkins, PA, USA 25–27 (2007).
211. Domingo E: Mechanisms of viral emergence. *Vet Res.* 41(6), 38 (2010).
212. Peters CJ: Emerging viral diseases. In: *Fields Virology*. Knipe DM, Howley PM (Eds). Wolters Kluwer, Lippincott Williams & Wilkins, PA, USA 605–625 (2007).
213. Griffin DE: Emergence and re-emergence of viral diseases of the central nervous system. *Prog. Neurobiol.* 91(2), 95–101 (2010).
214. Holland JJ, De La Torre JC, Steinhauer DA: RNA virus populations as quasispecies. *Current Topics Microbiol. Immunol.* 176, 1–20 (1992).

Dynamics of *In Vitro* Fitness Recovery of HIV-1[▽]

Ramón Lorenzo-Redondo, Antonio V. Bordería,[†] and Cecilio Lopez-Galindez*

Centro Nacional de Microbiología (CNM), Instituto de Salud Carlos III, Majadahonda, Madrid 28220, Spain

Received 11 June 2010/Accepted 16 November 2010

The study on the evolutionary consequences of an RNA viral population's fluctuations can be approached by *in vitro* experiments. This work describes the fitness recovery of HIV-1 after 20 large-population passages in 10 debilitated clones. The serial passages promoted an increase in viral fitness. In addition, we detected a significant number of mutations fixed in the complete genome consensus sequence of the final viral populations. Among the mutations, events of convergent evolution with important phenotypic characteristics occurred in several independent clones. One common change, V35I, in the nuclear localization signal of the p17 protein appeared in four viruses of three different lineages. Other common alterations mapped in position E196K of the reverse transcriptase or in position S316K of the V3 loop of the gp120 residue that is associated with the X4/R5 phenotype. Together with this mutational analysis, we studied the quasispecies heterogeneity of the initial and final viruses, revealing that fitness increase correlated with an augmentation in the genetic heterogeneity of viral quasispecies. However, while heterogeneity was mostly composed of synonymous (*dS*) mutations in the first 10 passages performed, at passage 21 it switched to nonsynonymous (*dN*) substitutions, with significant differences in *dN* – *dS* values between passages 11 and 21. In summary, the HIV-1 *in vitro* fitness recovery depicts a multiphase process occurring first by generation of mutations followed by fixation of the beneficial ones, depicting a classical Darwinian process.

Viral population fluctuations are recurrent in HIV-1 natural infections, and they have important consequences for viral evolution. During HIV-1 transmission, only one or a few variants are transmitted (29), producing a population bottleneck. Subsequent to this bottleneck, a great expansion of the viral population occurs in the infected individual which results in the fitness recovery of the virus (36). This increase in viral size and the high levels of polymerase errors result in the generation of a swarm of mutants, known as viral quasispecies (8). These quasispecies are the target of evolution (10). The analyses of the dynamics of the fitness recovery, as well as the study of the genetic variability generated and the mutations fixed in the genomes, are important for the understanding of HIV-1 pathogenesis.

RNA viruses, because of their high generation of genetic heterogeneity, rapid replication time, and large population sizes, constitute an ideal model to study evolution (7, 9, 17). Several RNA viruses, mainly vesicular stomatitis virus (VSV), foot-and-mouth disease virus (FMDV), or $\Phi 6$, have been used in *in vitro* works to study viral evolution after serial passages (26, 27), plaque to plaque (Muller's ratchet) (3, 4, 25), in large or alternating population sizes (11). Moreover, quasispecies heterogeneity has been found to be implicated in viral pathogenesis in poliovirus (37), mumps (31), West Nile virus (19), and HIV (1).

To help in the understanding of the variation occurring *in vivo*, we have performed *in vitro* evolution experiments with HIV-1. In these experiments, biological clones obtained from a patient's viral isolate (30) were subjected to serial bottleneck passages which produced drastic fitness losses (42) with an

unusual distribution of associated mutations (41). Subsequent large-population passages of these debilitated clones showed a rapid fitness increase in only 10 passages (1, 40). We disclosed that this recovery was found to be associated with an increase in viral quasispecies heterogeneity and occasionally with an invariable consensus sequence (1).

In the present work, we have extended, until passage 21, the large-population recovery passages in 10 HIV-1 clones, two of which were passaged in duplicate (1). At passage 21, there was an increase in the global fitness and in the genetic heterogeneity of the clones. We also obtained, as previously found in passage 11, a correlation between heterogeneity and fitness. In contrast with what occurred in passage 11, at passage 21 viral populations accumulated higher levels of nonsynonymous mutations in the quasispecies. We identified in the global sequences mutations with important phenotypic traits that arose in the absence of the selective pressure of the immune system. The overall process of the *in vitro* fitness recovery in HIV-1, which started by an initial phase of generation of variability and was followed by the selection of the beneficial changes in the populations, can be described as a Darwinian process.

MATERIALS AND METHODS

Cells, viruses, and biological cloning. Isolate s61, obtained in 1989 by standard coculture procedures from a perinatally infected 4-year-old boy in Madrid, Spain (30), was the origin of the HIV-1 parental population. From this viral population, six biological clones (D, E, G, H, I, and K) were derived by plating on MT-4 cells (16). In previous works, these clones were subjected to up to 15 serial plaque-to-plaque transfers, and populations D15, I15, and K15 were obtained. Clones E11, G7, and H12 became extinct before the 15th passage (42) (Fig. 1). The effects observed and the appearance of mutations in these bottleneck passages were described in references 41 and 42. For further studies, from these final plaque populations two clones were obtained from lineages D, G, I, and K and one was obtained from lineages E and H.

Large-population passages of viruses. The resulting 10 clones, D1, D2, E1, G1, G2, H1, I1, I5, K1, and K2, which are the initial populations of the present study, were subjected to 10 large-population passages by infecting 2.5×10^6 or

* Corresponding author. Mailing address: Centro Nacional de Microbiología, Instituto de Salud Carlos III, Carretera de Pozuelo Km 2, Majadahonda, 28220 Madrid, Spain. Phone: 34 91 822 39 18. Fax: 34 91 509 79 19. E-mail: clopez@isciii.es.

[†] Present address: Institut Pasteur, Department of Virology, Viral Populations and Pathogenesis Group, Paris 75724 Cedex 15, France.

[▽] Published ahead of print on 24 November 2010.

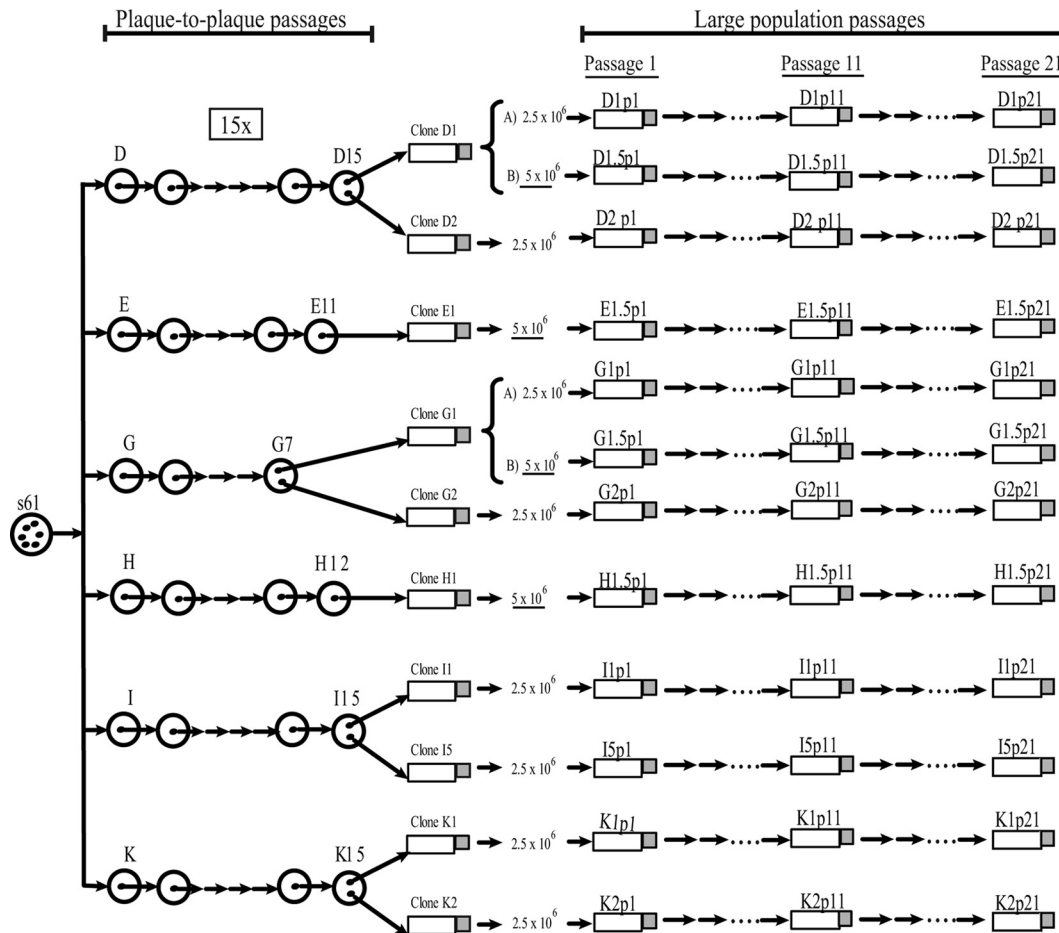


FIG. 1. Genealogy of the viral clones studied. Representation of the serial plaque-to-plaque transfers (circles in the left part of the figure) performed in previous works (41, 42). One or two clones were obtained from the final populations after the plaque-to-plaque transfers and were designated D1, D2, E1, G1, G2, H1, I1, I5, K1, and K2. Large-population passages (10 and 20) with these clones (rectangles, arrows, and dots; rectangles represent culture flasks) were performed in 2.5×10^6 and 5×10^6 MT-4 cells. The experimental procedures and the origin of the natural isolate s61 are described in references 30 and 42 and in Materials and Methods. Viral populations are indicated by letters that identify the clone used, followed by p1 for the initial population, p11 for passage 11, and p21 for passage 21. For clones D1, E1, G1, and H1, the number 5 after the period indicates that these clones were passed in 5×10^6 MT-4 cells. Clones D1 and G1 that are grouped by keys were passed in parallel in 2.5×10^6 cells (designated A) and in 5×10^6 MT-4 cells (designated B).

5×10^6 MT-4 cells (Fig. 1). The titers of the viruses were determined in MT-2 cells and expressed as the 50% tissue culture infective dose (TCID₅₀) per milliliter. Infections were performed, to minimize potential recombination events, at a low multiplicity of infection (MOI) of about 0.01 TCID₅₀/ml. Clones D1 and G1 were passed in parallel in 2.5×10^6 and 5×10^6 MT-4 cells (designated A and B, respectively, in Fig. 1). The results and characterization of the first 10 passages are summarized in reference 1. In the present work, using the same method, we performed the cultures until passage 21 (Fig. 1). Viruses were recovered from the culture supernatant when cytopathology was complete (5 to 7 days postinfection) and used to initiate the following passage, using fresh MT-4 cells (30). Cells were cultured in closed bottles, and the possibility of cross-contamination was monitored with uninfected control cultures. Enzyme-linked immunosorbent assay (ELISA) protein p24 determination was carried out in the culture supernatant with an Elecsys 2010 analyzer using a p24 detection kit (Roche Diagnostics, HIV Ag.HIV [groups M and O] p24 antigen).

DNA extraction and complete genome sequencing. For the determination of the consensus complete genome sequence, viral DNA was extracted from the cellular pellets by a standard phenol-chloroform method. PCRs used consisted of external and internal amplifications, using the Expand high-fidelity enzyme (Roche) with six different pairs of oligonucleotide primers that cover the entire HIV-1 genome. The sequences of the primers used in the amplifications, spanning the 5'LTR-gag (where LTR stands for long terminal repeat) (1,880 bp), pol-vpr (3,450 bp), and vpr-3'LTR (3,923 bp) regions, are described in reference

1. Both external and internal amplifications involved 35 cycles with temperatures chosen according to primer sequence. For primer location and numbering of nucleotide changes, refer to isolate HIV-1 HXB2 in the Los Alamos database (20). Sequences were determined on the two DNA strands with an ABI 3700 automatic sequencer. Multiple sequence alignments were performed using the CLUSTAL W program (35).

Fitness assay. Relative fitness values of the viruses obtained by large-population passages were determined by growth competition experiments as previously described (17). Briefly, 10^5 MT-4 cells were coinfecting with 10^3 TCID₅₀/ml of the virus to be tested mixed at a 1:1 ratio with the reference clone (J1). Coinfections were performed in triplicate for a maximum of 5 passages. The proportion of the two viruses in each passage was quantified by GeneScan analysis. J1 contains a deletion of 45 nucleotides in the V1/V2 loop of the *env* gene, which allows the differentiation of the two viruses in the GeneScan analysis. The fitness values per competitive transfer series were calculated from the slope of the exponential plot, as previously described (17). The statistical significance of fitness differences between populations in each passage was tested using a slope comparison test implemented in the GraphPad software. Although fitness estimations give relative fitness values, the use of the same reference strain allows the comparison of values between viruses.

GeneScan analysis. For fitness determination, to avoid recombination during the reverse transcriptase reaction when analyzing viruses from the culture supernatant, quantification of the viruses was performed on proviral DNA. Prior to this decision, a comparison between estimations performed on proviral DNA or

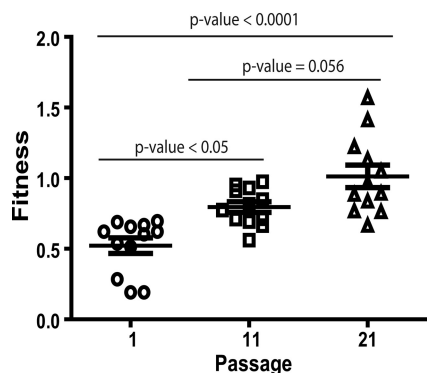


FIG. 2. Fitness increases throughout serial passages of HIV-1 clones. Representation of fitness values of each virus at passage 1 (circles), passage 11 (squares), and passage 21 (triangles). Mean values per passage are shown with their standard errors. Means were compared using a one-way ANOVA with a repeated-measures test followed by a Bonferroni's correction for multiple comparisons. *P* values of each comparison appear above the groups compared.

from viral RNA from the culture supernatant showed no statistical differences. The region encoding the V1/V2 loops of gp120 of the competing viruses, which were different in length, was amplified by PCR using one of the primers labeled with the fluorescent dye 6-carboxyfluorescein (6-FAM). The primers used for the V1/V2 amplification were 34 (5'-6FAM-GTCACAGTCTATTATGGGGTACC TGTGT-3') and 31 (5'-ACCTCAGTCATTACACAGGCCTGCAGCGC-3'). The PCR products were diluted 2- to 10-fold in water to obtain peak intensities in the linear range, and 1 μ l of these dilutions was mixed with 9 μ l of 1000 ROX size standard (Applied Biosystems) previously diluted 10-fold with formamide. Samples were run on an ABI Prism 3700 machine, and data were analyzed using GeneMapper version 4.0. Peak areas were used to calculate the proportion of each virus versus the reference virus J1 and to derive fitness.

Quasispecies analysis. For the analysis of the mutant spectra, we selected four genomic regions that were amplified by using three primer pairs. The first pair amplified DNA positions 672 to 1363 of HXB2 (692 nucleotides) that spanned the leader sequence and the p17 protein in the *gag* gene. The second pair amplified DNA positions 6045 to 6734 (690 nucleotides) encompassing *vpu* and the V1/V2 region of *env*. This fragment was divided into two regions, one corresponding to the *vpu* gene and the other to the V1/V2 *env* gene. The third pair amplified DNA positions 7039 to 7641 (603 nucleotides) that comprised the V3/V4 region of *env*. In total, the three amplifications covered 2,005 genomic residues, or approximately 20% of the HIV-1 genome. The PCR products were cloned using TA TOPO cloning (Invitrogen), following the manufacturer's instructions. At least 20 independent clones were used to represent the viral quasispecies of every population in each passage. Nucleotide sequences from these clones were used for quasispecies characterization and for the quantification of genetic heterogeneity.

Genetic distances within quasispecies calculation. In each region, jModelTest (28) was used to estimate the best model that fit the evolution of the sequences and to obtain the model average estimates of each evolutive parameter. There was no difference between regions in the evolutionary model best adapted to the data, which was in all cases the general time-reversible model with gamma distribution of mutations (GTR+G). Afterwards, genetic distances within quasispecies were calculated using the maximum likelihood method with the model average estimates of the parameters given by jModelTest. This calculation was performed with the PAUP software (33). As no significant differences in genetic distances were observed between gene regions within each virus and passage (repeated-measures analysis of variance [ANOVA] *P* value of >0.05), mean values in the four regions were used for the global analysis per passage.

Analyses of synonymous and nonsynonymous mutations within quasispecies in each region compared with the initial passage were performed by applying the modified Nei-Gojobori method (24) in the MEGA version 4.0 program (34). Values of nonsynonymous substitutions per nonsynonymous site (*dN*) and synonymous substitutions per synonymous site (*dS*) in each virus were calculated using the Jukes-Cantor model, and mean *dN* and *dS* values for each passage were tested for statistical differences.

Statistical analyses. Data were analyzed using GraphPad Prism and PasW Statistics 18 software. For comparisons of values between passages, one-way

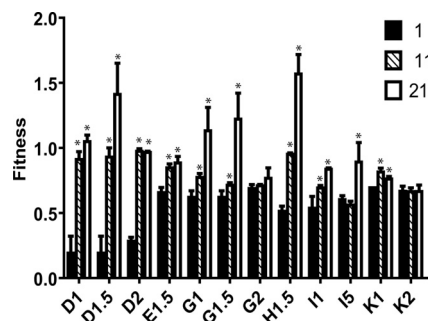


FIG. 3. Fitness values per virus and passage. Biological fitness values obtained by competition assays as explained in Materials and Methods. Columns represent each virus value in every passage (the patterns association with each passage are shown in the legend). *, the difference with the initial passage is statistically significant ($P < 0.1$) using a slope comparison test.

ANOVA with repeated measures and Bonferroni's correction for multiple comparisons were used. Spearman correlation analyses were performed to study the correlation between fitness and overall heterogeneity and within passages. In addition, linear mixed-effect models were tested using maximum likelihood, taking into account the passages and random effects within viruses.

SplitsTree quasispecies networks. Clone sequences of the quasispecies were used to perform networks in the different viral lineages using SplitsTree4 software (18). The median network method was used in order to visualize the dynamics of the quasispecies in every region of each virus, using the S61 original virus as an outgroup. Distances were measured as the number of differences.

Nucleotide sequence accession numbers. Sequences have been deposited in GenBank with the accession numbers HM 469972 to 469983.

RESULTS

This paper describes the process of HIV-1 fitness recovery of 10 previously debilitated viral clones after large-population passages. The results of the recovery of the same clones after the initial 10 passages have been presented previously (1). This work summarizes and compares the results after 10 additional passages, until passage 21.

HIV-1 fitness increases after 20 large-population passages. We carried out the fitness recovery of viral clones after 20 serial passages in MT-4 cells. Fitness of the initial populations at passage 1 was 0.52 ± 0.05 (mean \pm standard error) and 0.79 ± 0.04 at passage 11, and in passage 21, mean fitness had increased to 1.03 ± 0.09 . Using a one-way ANOVA with a repeated-measures test, followed by a Bonferroni's correction for multiple comparisons, the difference between passages 1 and 11 was significant at *P* values of <0.05, while between passages 1 and 21 it was even more significant, with *P* values of <0.0001 (Fig. 2). The difference between passages 11 and 21 was marginally significant, with a *P* value of 0.056. In addition to the global fitness increase throughout the passages, we observed a wider dispersion of the fitness values. Overall, these results showed that *in vitro* serial recovery passages produced an increase in the global fitness of the viruses.

Apart from the global fitness increase, we also examined individual viruses, and we detected, as shown in Fig. 3, a very heterogeneous behavior. Five viruses (D1.5, G1, G1.5, H1.5, and I1) that increased their fitness at passage 11 continued improving at passage 21. Four viruses (D1, D2, E1.5, and K1) with a previous increase became stable at passage 21, one virus without previous increase was now increasing (I5), and two viruses (G2 and K2)

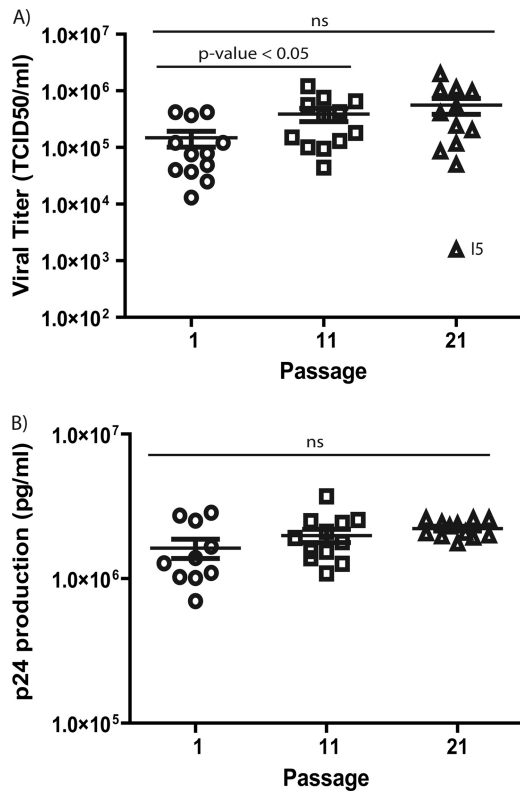


FIG. 4. Viral titers and p24 protein production during serial passages. (A) Viral titers of viruses at passages 1, 11, and 21. Titers were obtained by infecting MT-2 cells with the corresponding supernatants and expressed as TCID₅₀ (see Materials and Methods). To compare the mean values, a one-way ANOVA with a repeated-measures test and a Bonferroni's correction for multiple comparisons were used, and the *P* value is given above the data points. ns, not significant. (B) p24 protein was quantified using an indirect ELISA, as described in Materials and Methods. Mean values of passages 1, 11, and 21 are shown in pg/ml and were used to compare passages. Comparisons between passages were not significantly different using the same statistical tests. Symbols are as described in the legend to Fig. 2.

were still unable to recover their fitness at passage 21. In summary, 10 of the 12 viruses had increased their fitness throughout the passages; however, the different patterns of fitness recovery in each virus explained the dispersion of the global data through the passages and suggested different recovery processes.

Viral quantification during passages. To explore the recovery process of viruses at passage 21, we estimated their infectivity by measuring the viral titers in TCID₅₀ and viral production by p24 protein quantification. With the exception of clone I5, which showed a decay in viral titer (Fig. 4A), viral infectivity had not increased significantly compared with that of passage 11.

Protein p24 viral production remained constant until passage 21. However, a homogenization in values at passage 21 was observed which is illustrated by the means of the standard error (SEM), which were 2.5×10^5 pg/ml in passage 1, 2.1×10^5 pg/ml in passage 11, and 0.8×10^5 pg/ml in passage 21 (Fig. 4B). These results point to an optimization of the viral production indicated by the increase in p24 in the initial cultures and the stability in viral titers through the serial passages.

Emergence and fixation of mutations in the complete genome consensus sequences at passage 21. To study the potential causes of the fitness increase in viral clones, an analysis of mutations emerging during passages was carried out in the complete genome at the consensus sequence level. Mutations emerging during passages increased very significantly, from 24 mutations in passage 11 (1) to 95 mutations in passage 21 (including 16 alterations in common with passage 11) (Fig. 5 and Table 1). More than 80% of the changes were nonsynonymous (Table 1). Of the 24 mutations fixed in passage 11, 16 (67%) remained fixed at passage 21. Interestingly, the mutations that were not fixed reverted to the initial nucleotide before passages, showing the reemergence of these preexisting variants (data not shown). Thus, this result indicated the importance of mutations fixed at the initial passages for fitness recovery. In addition, the fitness increase observed in passage 21 was associated with a very significant augmentation of the mutations fixed in the complete genome consensus sequences.

Moreover, five events of convergence between viruses were observed (Fig. 5). Change 1 in Fig. 5 corresponds to one of the changes fixed in the LTR from a previous work (40). Another convergence event, change V35I in p17 protein (change 2 in Fig. 5), appeared in four viruses of three different lineages (G1.5, G2, I5, and K2 clones). It is important to highlight that these viruses had not accumulated common mutations during the bottleneck passages (41). Thus, no compensating effect to a previous mutation can be attributed to the V35I mutation, and probably it appeared because of its beneficial effect on fitness. This position has been associated with fitness recovery of protease inhibitor resistance mutants (15). This change is located in the C end of an α -helix corresponding to the nuclear localization signal (2). Another change in the reverse transcriptase (RT) in the *pol* gene product in position E196K (number 3 in Fig. 5) appearing in G clones could have important effects due to its proximity to part of the enzyme catalytic site in residues 185 and 186. The S316K change (number 4 in Fig. 5) that was fixed in the V3 loop in gp160 in two G clones occurred precisely in one of the positions associated with the X4/R5 phenotypic change and with the same amino acid change (Fig. 5). Change 5 in Fig. 5, corresponding to position L16F in p7, has not been found in the literature to be associated with significant effects. It is important to point out that all these changes appeared without any external selective force operating, such as immune system or antiviral treatment. The elevated number of mutations fixed and the convergence events observed suggested a strong positive selection operating in the fitness recovery at passage 21.

Detection of deletions in the V1/V2 loop in the *env* gene. All viruses, except viruses E1.5 and H1.5, had an initial length identical to that of the primary isolate (41) of 218 nucleotides in the V1/V2 loop in the *env* gene. Along with the emergence and fixation of mutations described, a majority of viruses (9 out of the 12 viruses passaged) (Fig. 1) acquired deletions in this region during the passages. These deletions showed diverse sizes, from 54 nucleotides in clone D2 to 6 nucleotides in K2. Some viruses presented only one dominant deleted variant (E1.5, H1.5), while others had up to six variants (K2) (Fig. 6). Once they appeared in the population, all these deleted variants, except clone K2, became dominant in the quasispecies (Fig. 6). The appearance and dominance of these deleted forms in the quasispecies support their importance for the *in vitro* viral fitness improvement.

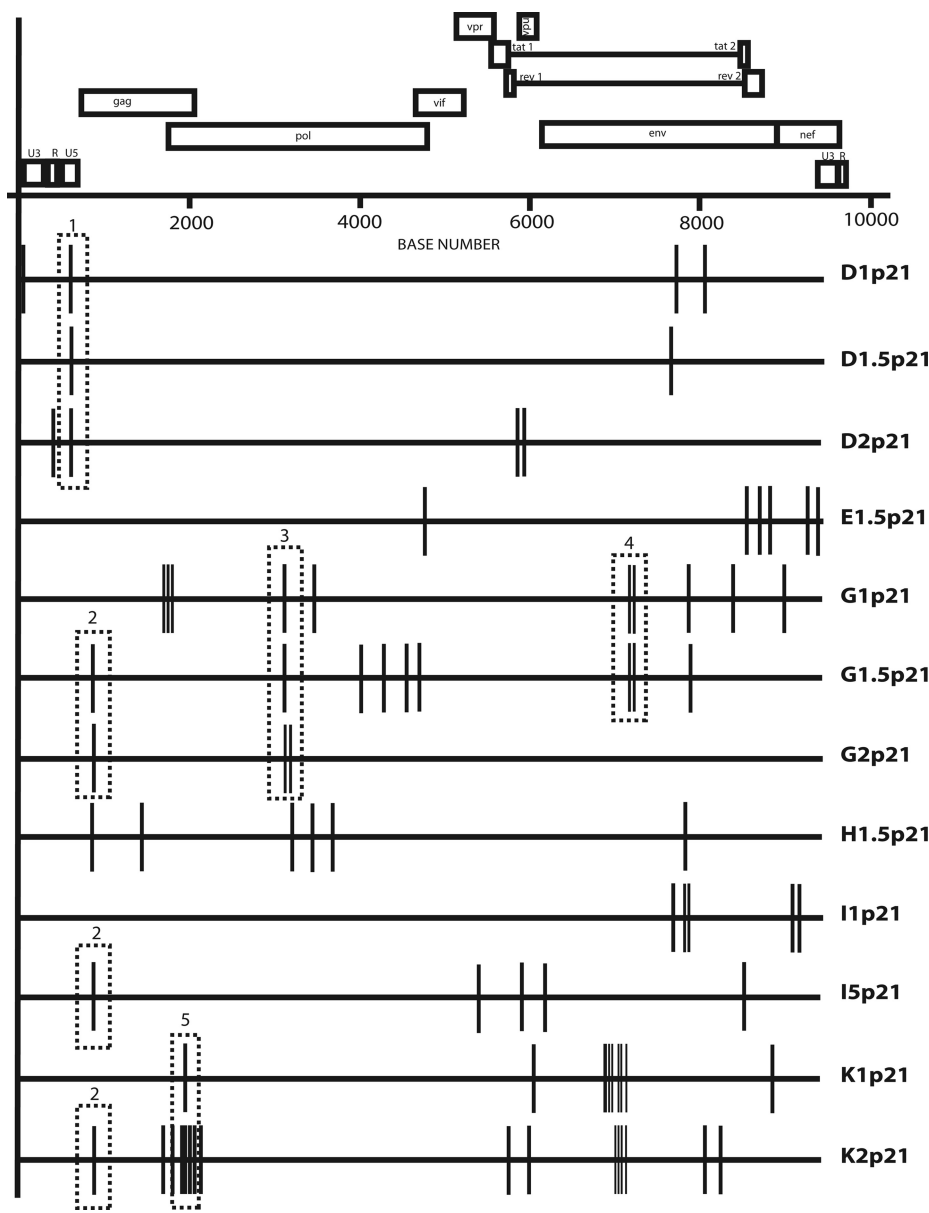


FIG. 5. Mutations fixed in the consensus complete sequence from viral populations at passage 21. Schematic representation based on the highlighter tool (<http://www.hiv.lanl.gov/content/sequence/HIGHLIGHT/highlighter.html>) of the mutations fixed in the complete genome consensus sequence in each virus studied at passage 21. Vertical lines indicate changes compared with the initial virus. Rectangles with dashed borders specify events of convergence numbered as follows: 1, G593T (in LTR); 2, V35I (in p17); 3, E196K (in RT) (G2p21 virus acquired this amino acid change through two nucleotide changes); 4, S316K (in gp120) (G1p21 and G2p21 viruses acquired this amino acid change through two nucleotide changes); 5, L16F (in p7).

TABLE 1. Synonymous and nonsynonymous mutations fixed at passage 21^a

Type of mutation	New mutations at passage 21			Mutations fixed from passage 11			Total
	No.	% with NC changes	% without NC changes	No.	% with NC changes	% without NC changes	
All	79	100		16	100		95
Nonsynonymous	64	81	83.1	11	68.8	84.6	
Synonymous	13	16.5	16.9	2	12.5	15.4	
Noncoding (NC)	2	2.5		3	18.8		

^a Number and percentage of synonymous and nonsynonymous changes within the new mutations fixed at passage 21 and changes that remained fixed from passage 11. Mutations in noncoding regions are also indicated. Percentages are calculated both with these NC changes and without them.

Quasispecies heterogeneity increases throughout the recovery passages. In addition to the complete genome sequence performed at the global level, a quasispecies analysis was carried out in four genomic regions. To study quasispecies genetic heterogeneity, mean distances within quasispecies obtained by maximum likelihood, as explained in Materials and Methods, were compared between passages. Mean genetic distances increased from $3.234 \times 10^{-3} \pm 0.202 \times 10^{-3}$ substitutions per site in passage 1 to $3.725 \times 10^{-3} \pm 0.279 \times 10^{-3}$ substitutions per site in passage 11 and $4.345 \times 10^{-3} \pm 0.343 \times 10^{-3}$ substitutions per site in passage 21 (Fig. 7). These differences, however, were only statistically significant between initial and final passages when

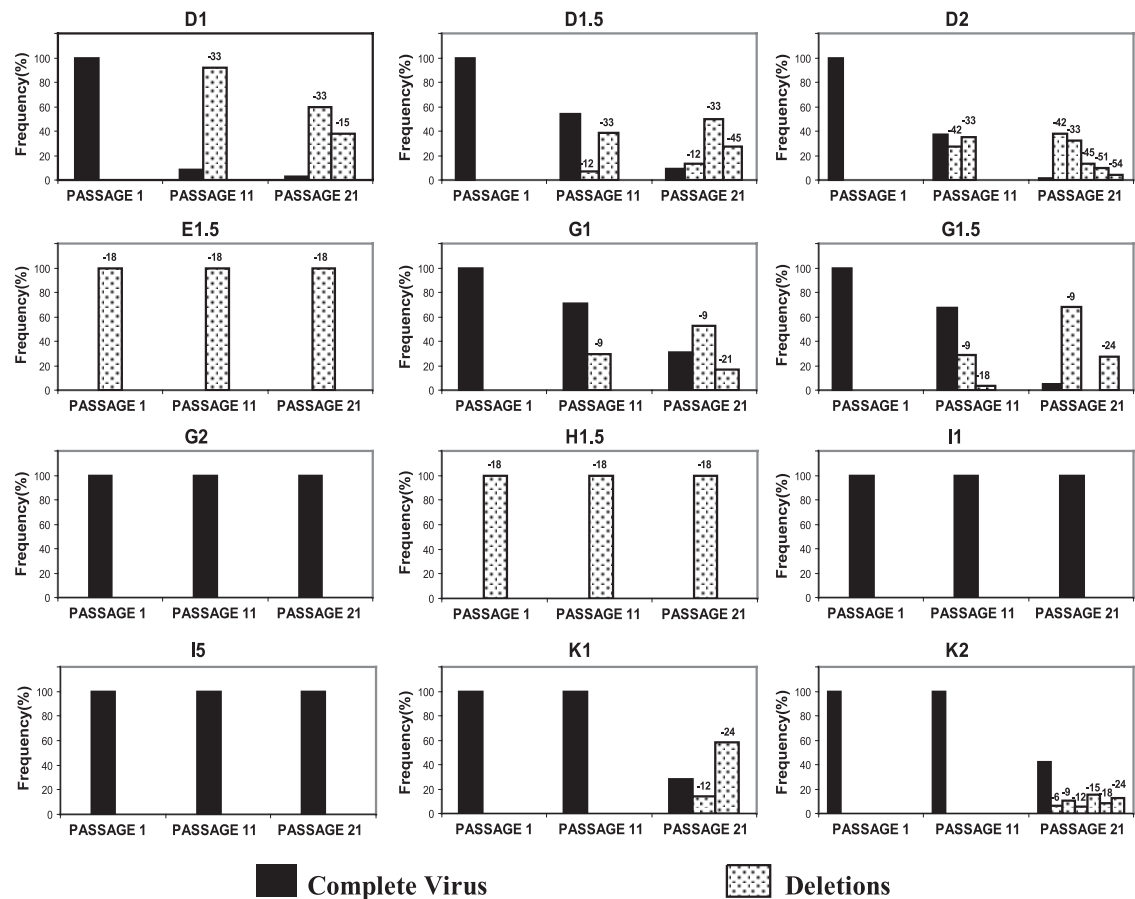


FIG. 6. Deletions in the V1/V2 region of the *env* gene of viruses. Emergence of genomes with deletions in the (V1/V2) region of *env* is represented for passages 11 and 21, expressing their frequency in the quasiespecies. Nondeleted genomes are represented by filled bars, while deleted genomes are shown with white bars with black dots. The length of the deletion is indicated above its corresponding bar.

tested using one-way ANOVA with the repeated-measures test followed by a Bonferroni's correction for multiple comparisons (P value of <0.05). We also observed an increase in the dispersion of data at passage 21, as can be observed in the SEM values. In

summary, quasiespecies heterogeneity has been increasingly generated throughout the large-population passages.

Significant global correlation between heterogeneity and fitness. Since in the initial passages of fitness recovery we found a significant correlation between quasiespecies heterogeneity and fitness (1), we carried out the same analysis at passage 21 (Fig. 8). We performed a Spearman correlation test among all clones studied, and a global significant correlation between fitness and heterogeneity ($P < 0.0309$) was obtained (Fig. 8A). This result demonstrated the importance of the generation of variability for the HIV-1 *in vitro* fitness recovery.

However, when we performed the same statistical test within passages, we detected a multiphase process (Fig. 8B). In the initial populations, we observed low heterogeneity and fitness, while in passage 11 a tendency of increasing heterogeneity and fitness was attained that was not significant with all viruses but became very significant (P value of 0.0233) when not taking into account virus G2p11, an outlier, because of an accumulation of mutations in *vpu* (Fig. 8B). At passage 21, this correlation was not significant due to three viruses (D1.5p21, G1.5p21, H1.5p21) that showed higher fitness values but decreased heterogeneity (see Fig. 8A). In all the remaining viruses, however, there was a general pattern of augmentation of heterogeneity and fitness through passages. The existence of these differences between passages led us to confirm the over-

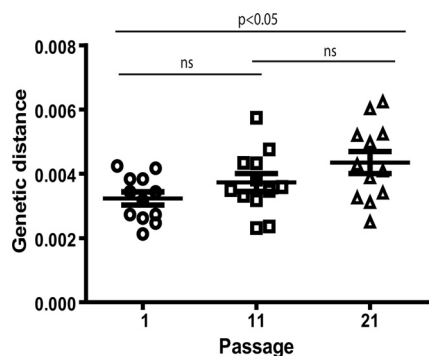
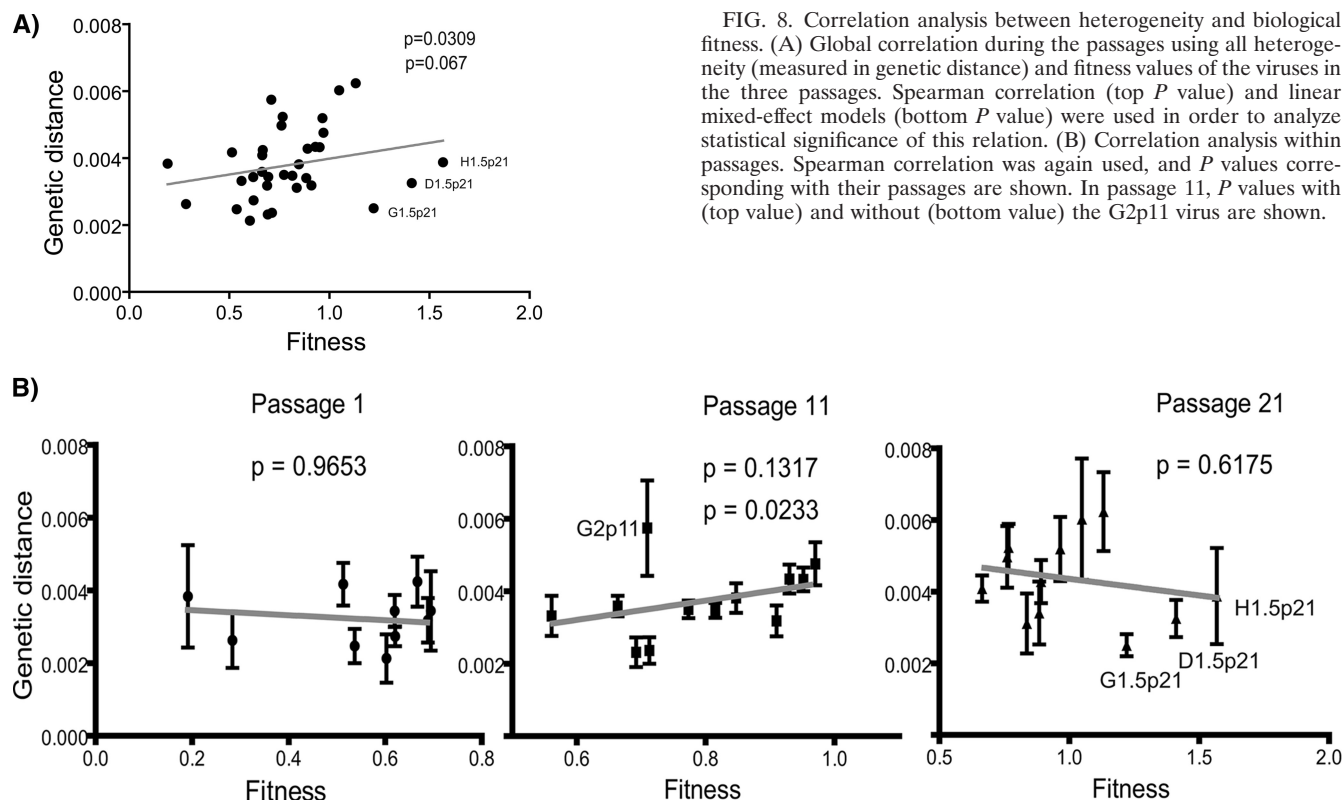


FIG. 7. Quasiespecies heterogeneity throughout the passages. Heterogeneity was measured as the mean genetic distance (number of substitutions per site) as explained in Materials and Methods. Quasiespecies heterogeneity was analyzed using mean values of each virus in the four regions examined. Comparison between passages was performed, and P values (Bonferroni corrected) are shown above each pair of columns. ns, nonstatistically significant difference. The values, means, and standard errors per passage for viruses are shown using symbols as described in the legend to Fig. 2.



all correlation previously obtained, taking into account the influence of passages. For this purpose, we tested linear mixed-effect models considering the passages as repeated measures, obtaining a correlation between heterogeneity and fitness, with a P value of 0.067 in the best model obtained by Akaike Information Criterion. The selected model included an unstructured covariance matrix of repeated effects and considered random effects within viruses. This result lowers the overall relationship between heterogeneity and fitness due to the variability of these factors throughout the passages. The different phases observed in the fitness recovery depict a dynamic process and point to the operation of different selective pressures during HIV-1 *in vitro* fitness recovery.

Analysis of selective forces operating in the recovery process. To confirm this hypothesis, we performed an analysis of synonymous and nonsynonymous changes within quasiespecies at different passages. Mean synonymous substitutions per synonymous site (dS) and mean nonsynonymous substitutions per nonsynonymous site (dN) values in each virus (as explained in Materials and Methods) were grouped per passage and compared using a one-way ANOVA with a repeated-measures test, followed by a Bonferroni's correction for multiple comparisons. As observed in Fig. 9, heterogeneity of viruses changed its composition throughout the passages. While in the initial passages heterogeneity was contributed mainly by synonymous mutations, in the final viruses it switched to nonsynonymous substitutions. When $dN - dS$ values were compared, we found significant differences between passages 11 ($-6.5 \times 10^{-4} \pm 2.9 \times 10^{-4}$) and 21 ($4.6 \times 10^{-4} \pm 3.9 \times 10^{-4}$) (P value < 0.05). This analysis supports that the initial steps of the fitness

recovery were dominated by purifying selection, whereas positive selection drove the evolution in the subsequent passages.

DISCUSSION

In vitro studies provide very useful information to understand natural viral infections. This work describes fitness recovery in debilitated HIV viral clones after 20 large-population passages. At this passage, viruses had improved viral fitness, associated with an increase of the genetic heterogeneity of the viral quasiespecies and with the fixation in the global complete

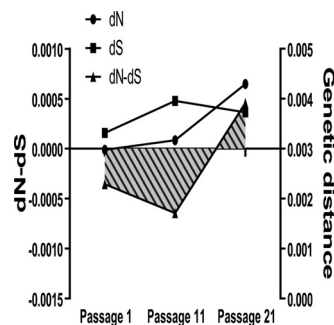


FIG. 9. $dN - dS$ analysis of the recovery passages. Synonymous and nonsynonymous mutations were calculated according to the Nei-Gojobori method. dN and dS were expressed using Jukes-Cantor model and are given by the number of synonymous and nonsynonymous differences from the possible number of synonymous and nonsynonymous sites. Mean $dN - dS$ values are shown as filled areas, and stripes represent negative values.

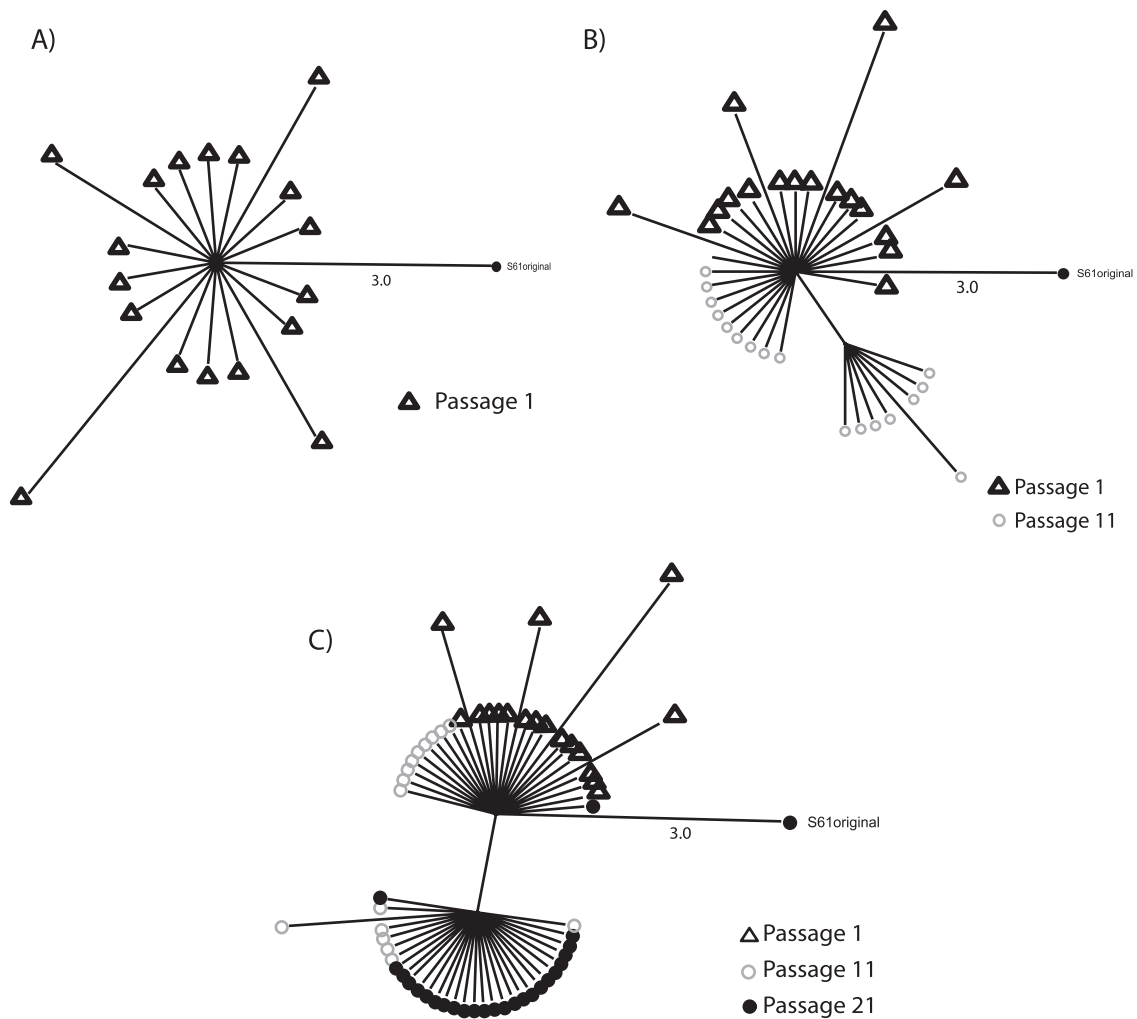


FIG. 10. Network analysis of quasispecies in LTR-*gag* of clone H1.5. Quasispecies sequences in the LTR-*gag* fragment for H1.5 virus in each passage were added sequentially in order to observe their evolution with time. SplitsTree4 was used to perform this analysis using the median network method. Distances between nodes were measured in numbers of mutations. (A) Sequences from passage 1; (B) sequences from passage 1 plus passage 11; (C) sequences from passage 1 plus sequences from passage 11 and passage 21. As a reference, distance between the outgroup S61 original virus and the central node is indicated.

genome sequence of multiple mutations, some providing important phenotypic characteristics.

To identify the specific causes of the fitness increase, we analyzed the complete genomic sequence of the viruses before and after the passages. In contrast to what was observed at passage 11, in which only 24 mutations were fixed among the clones, in passage 21, 79 new mutations were observed. Of interest, 16 (67%) of the mutations fixed at passage 11 were maintained at passage 21, suggesting that these mutations constituted the basis from which viral recovery started (Table 1).

The difference in fitness among clones could also be attributed to the appearance of different deletions in the V1/V2 loop (Fig. 6). The V1/V2 region of viruses at passage 1, except in E and H clones, was 218 nucleotides in length. The original s61 virus, from which all these clones were derived, also had 218 nucleotides. This is a regular length for primary isolates since the average size of this region in the Los Alamos database is about 210 nucleotides. The importance of these deletions is

given by their frequency and by their imposition in the viral populations after the passages. In fact, only three of the viruses maintained the full-length loop at passage 21, in contrast with five viruses at passage 11. The imposition of the deleted genomes is not associated with a single deleted form, which only occurred in viruses E and H, but rather to a set of different deleted variants with up to six forms (viruses D2 and K2). It has been shown that viruses of subtype C with shorter V1/V2 regions are better transmitted and easier to neutralize than average (6).

Looking at the dynamics of the fitness recovery, we can describe HIV-1 fitness recovery as a multiphase process. At the selective level, what was important was the change in the composition of the mutations appearing in viral quasispecies, measured as $dN - dS$ and shown in Fig. 9. While at passage 11 there was a dominance of synonymous changes, this pattern shifted at passage 21 to a dominance of nonsynonymous alterations, pointing to the operation of selective pressures. We

used the $dN - dS$ relationship instead of the most commonly used dN/dS ratio, because we were dealing with quasispecies sequences that are very similar. In fact, in this set of sequences, 6% of them had a zero value for dS , and as the ratio tends to infinity, these data cannot be used. However, to confirm this result, we calculated the dN/dS ratio, and the same significant difference between passages 11 and 21 with a P value of <0.05 was obtained.

To further study the dynamics of this process, we analyzed the quasispecies nucleotide sequences in a network analysis. We used this method with the sequences of the clones in each region because it allowed an excellent visualization of the viral quasispecies and its dynamics through the passages. This analysis is shown in Fig. 10 for virus H1.5 in the LTR-*gag* region as a representative example of other viruses and regions. H1.5 virus generated variability in the initial passages (Fig. 10A), and when a beneficial mutation arose, the A95G change in p17 (Fig. 10B), the population moved to this new point, with only a minor representation of the previous mutants (Fig. 10C). From this new point, the virus continued to generate variability, and the majority of the variants at passage 21 were in the new position (Fig. 10C). We observed in every lineage, at least in one of the regions studied, a similar pattern of evolution. All these data support the notion, as previously described (13), that the HIV-1 *in vitro* fitness recovery follows a Darwinian model.

The observed variability in fitness values of individual clones (Fig. 3) indicated that each virus was pursuing a different evolutive pathway. This difference was also sustained by the different mutations that each virus was incorporating (Fig. 5). However, the events of convergence found in some viruses showed the existence of common pathways in the process of recovery (Fig. 5). Of special interest were the mutations identifying the convergent events, such as V35I in p17, mutation S136K in the V3 loop of gp120 (5), and mutation E196K in the RT. It is essential to highlight that these mutations that have also been associated with important phenotypic changes were fixed in the quasispecies under *in vitro* conditions without any external selective pressure. These findings make clear that some phenotypic alterations, like antiretroviral resistance mutations (22, 23) among other properties, can preexist in viral populations or can be naturally generated.

We previously demonstrated that among the causes of the fitness recovery is the generation of heterogeneity in the viral quasispecies (1). This is in agreement with studies performed in other viruses (27). The importance of quasispecies diversity in pathogenesis has been demonstrated in poliovirus, where high-fidelity viral strains produced less heterogeneous and less pathogenic viral populations (37); also, the pathogenesis of a mumps vaccine strain has been related to the quasispecies heterogeneity (31). In addition, a mutant polio strain which generates a limited heterogeneity is under study as a new poliovirus vaccine candidate (38).

The role heterogeneity plays in fitness recovery, apart from supporting the theoretical concept of quasispecies, also has connections to *in vivo* situations. Genetic analysis of a patient's viral population has been extensively used to gain information on the *in vivo* HIV-1 evolution but also has been related to HIV-1 pathogenesis (13, 32, 39). Primary isolates obtained sequentially along the infection time gained fitness at a signif-

icant rate that directly correlated with viral load and HIV-1 *env* C2V3 diversity (36). In addition, genetic variability contributed to an optimization of the virus and of the viral envelopes (21). The role that HIV-1 genetic heterogeneity has on *in vivo* pathogenesis can also be explained by the role of genetic heterogeneity in fitness increase as demonstrated in our *in vitro* experiments (1). In this study, we verified that the fitness recovery has a crucial initial step of variation generation. This situation has a clear parallel to *in vivo* situations where the control of genetic variation is considered one of the most important targets for antiviral treatment and vaccine studies (14), in which the inclusion of different combinations of immunogens is under investigation (12).

This *in vitro* study has provided insightful information to understand the fitness recovery in HIV-1 clones after serial passages. The viruses had improved viral fitness after increasing their genetic heterogeneity and the fixation of multiple mutations and deletions. We described different factors that have to be taken into account, and we highlighted the importance of the generation of variability to fuel selection after a transmission bottleneck. This study depicts a multiphase Darwinian process of *in vitro* fitness recovery of HIV-1.

ACKNOWLEDGMENTS

The helpful comments and suggestions of two anonymous reviewers are greatly appreciated, as well as the critical reading of the manuscript by Concepción Casado and Monica Gutierrez-Rivas.

Work at CNM is supported by grant SAF 2007-61036 of the Ministerio de Ciencia e Innovación, by Fundacion para la Investigacion y Prevencion del SIDA en España (FIPSE) grants 36558/06, 36641/07, 36779/08, and 360766/09, and in part by the Red Temática Cooperativa de Investigación en SIDA (Red de grupos 173) of the Fondo de Investigaciones Sanitarias de la Seguridad Social (FISs).

REFERENCES

1. Borderia, A. V., et al. 2010. Initial fitness recovery of HIV-1 is associated with quasispecies heterogeneity and can occur without modifications in the consensus sequence. *PLoS One* **5**:e10319.
2. Bukrinsky, M. I., et al. 1993. A nuclear localization signal within HIV-1 matrix protein that governs infection of non-dividing cells. *Nature* **365**:666–669.
3. Chao, L. 1990. Fitness of RNA virus decreased by Muller's ratchet. *Nature* **348**:454–455.
4. Clarke, D. K., et al. 1993. Genetic bottlenecks and population passages cause profound fitness differences in RNA viruses. *J. Virol.* **67**:222–228.
5. Cornelissen, M., E. Hogervorst, F. Zorgdrager, S. Hartman, and J. Goudsmit. 1995. Maintenance of syncytium-inducing phenotype of HIV type 1 is associated with positively charged residues in the HIV type 1 gp120 V2 domain without fixed positions, elongation, or relocated N-linked glycosylation sites. *AIDS Res. Hum. Retroviruses* **11**:1169–1175.
6. Derdeyn, C. A., et al. 2004. Envelope-constrained neutralization-sensitive HIV-1 after heterosexual transmission. *Science* **303**:2019–2022.
7. Domingo, E., C. Escarmis, L. Menendez-Arias, and J. J. Holland. 1999. Viral quasispecies and fitness variations, p. 141–161. *In* E. Domingo, R. Webster, and J. Holland (ed.), *Origin and evolution of viruses*. Academic Press, San Diego, CA.
8. Domingo, E., et al. 1985. The quasispecies (extremely heterogeneous) nature of viral RNA genome populations: biological relevance—a review. *Gene* **40**:1–8.
9. Domingo, E., C. R. Parrish, and J. J. Holland (ed.). 2008. *Origin and evolution of viruses*, 2nd ed. Elsevier, London, United Kingdom.
10. Eigen, M. 1971. Self-organization of matter and the evolution of biological macromolecules. *Naturwissenschaften* **58**:465–523.
11. Elena, S. F., et al. 1996. Evolution of fitness in experimental populations of vesicular stomatitis virus. *Genetics* **142**:673–679.
12. Fischer, W., et al. 2007. Polyvalent vaccines for optimal coverage of potential T-cell epitopes in global HIV-1 variants. *Nat. Med.* **13**:100–106.
13. Ganesan, S., R. Dickover, B. T. M. Korber, Y. J. Bryson, and S. M. Wolinsky. 1997. Human immunodeficiency virus type 1 genetic evolution in children with different rates of development of disease. *J. Virol.* **71**:663–677.
14. Gaschen, B., et al. 2002. Diversity considerations in HIV-1 vaccine selection. *Science* **296**:2354–2360.

15. **Gatanaga, H., et al.** 2002. Amino acid substitutions in Gag protein at non-cleavage sites are indispensable for the development of a high multitude of HIV-1 resistance against protease inhibitors. *J. Biol. Chem.* **277**:5952–5961.
16. **Harada, S., Y. Koyanagi, and N. Yamamoto.** 1985. Infection of HTLV-III/LAV in HTLV-I-carrying cells MT-2 and MT-4 and application in a plaque assay. *Science* **229**:563–566.
17. **Holland, J. J., J. C. de la Torre, D. K. Clarke, and E. Duarte.** 1991. Quantitation of relative fitness and great adaptability of clonal populations of RNA viruses. *J. Virol.* **65**:2960–2967.
18. **Huson, D. H., and D. Bryant.** 2006. Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* **23**:254–267.
19. **Jerzak, G. V., K. Bernard, L. D. Kramer, P. Y. Shi, and G. D. Ebel.** 2007. The West Nile virus mutant spectrum is host-dependant and a determinant of mortality in mice. *Virology* **360**:469–476.
20. **Kuiken, C., et al.** 2009. HIV sequence compendium 2009. Theoretical Biology and Biophysics, Los Alamos National Laboratory, Los Alamos, NM.
21. **Lassen, K. G., et al.** 2009. Elite suppressor-derived HIV-1 envelope glycoproteins exhibit reduced entry efficiency and kinetics. *PLoS Pathog.* **5**:e1000377.
22. **Najera, I., et al.** 1995. *pol* gene quasispecies of human immunodeficiency virus: mutations associated with drug resistance in virus from patients undergoing no drug therapy. *J. Virol.* **69**:23–31.
23. **Najera, I., et al.** 1994. Natural occurrence of drug resistance mutations in the reverse transcriptase of human immunodeficiency virus type 1 isolates. *AIDS Res. Hum. Retroviruses* **10**:1479–1488.
24. **Nei, M., and T. Gojobori.** 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.* **3**:418–426.
25. **Novella, I. S.** 2003. Contributions of vesicular stomatitis virus to the understanding of RNA virus evolution. *Curr. Opin. Microbiol.* **6**:399–405.
26. **Novella, I. S., et al.** 1995. Extreme fitness differences in mammalian and insect hosts after continuous replication of vesicular stomatitis virus in sandfly cells. *J. Virol.* **69**:6805–6809.
27. **Novella, I. S., et al.** 1995. Exponential increases of RNA virus fitness during large population transmissions. *Proc. Nat. Acad. Sci. U. S. A.* **92**:5841–5844.
28. **Posada, D.** 2008. jModelTest: phylogenetic model averaging. *Mol. Biol. Evol.* **25**:1253–1256.
29. **Salazar-Gonzalez, J. F., et al.** 2009. Genetic identity, biological phenotype, and evolutionary pathways of transmitted/founder viruses in acute and early HIV-1 infection. *J. Exp. Med.* **206**:1273–1289.
30. **Sanchez-Palomino, S., et al.** 1993. Dilute passage promotes expression of genetic and phenotypic variants of human immunodeficiency virus type 1 in cell culture. *J. Virol.* **67**:2938–2943.
31. **Sauder, C. J., et al.** 2006. Changes in mumps virus neurovirulence phenotype associated with quasispecies heterogeneity. *Virology* **350**:48–57.
32. **Shankarappa, R., et al.** 1999. Consistent viral evolutionary changes associated with the progression of human immunodeficiency virus type 1 infection. *J. Virol.* **73**:10489–10502.
33. **Swofford, D. L.** 1998. PAUP*: phylogenetic analysis using parsimony (* and other methods) version 4.0 beta. Sinauer, Sunderland, MA.
34. **Tamura, K., J. Dudley, M. Nei, and S. Kumar.** 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* **24**:1596–1599.
35. **Thompson, J. D., D. G. Higgins, and T. J. Gibson.** 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**:4673–4680.
36. **Troyer, R. M., et al.** 2005. Changes in human immunodeficiency virus type 1 fitness and genetic diversity during disease progression. *J. Virol.* **79**:9006–9018.
37. **Vignuzzi, M., J. K. Stone, J. J. Arnold, C. E. Cameron, and R. Andino.** 2006. Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population. *Nature* **439**:344–348.
38. **Vignuzzi, M., E. Wendt, and R. Andino.** 2008. Engineering attenuated virus vaccines by controlling replication fidelity. *Nat. Med.* **14**:154–161.
39. **Wolinsky, S. M., et al.** 1996. Adaptive evolution of human immunodeficiency virus-type 1 during the natural course of infection. *Science* **272**:537–542.
40. **Yuste, E., A. V. Borderia, E. Domingo, and C. Lopez-Galindez.** 2005. Few mutations in the 5' leader region mediate fitness recovery of debilitated human immunodeficiency type 1 viruses. *J. Virol.* **79**:5421–5427.
41. **Yuste, E., C. Lopez-Galindez, and E. Domingo.** 2000. Unusual distribution of mutations associated with serial bottleneck passages of human immunodeficiency virus type 1. *J. Virol.* **74**:9546–9552.
42. **Yuste, E., S. Sanchez-Palomino, C. Casado, E. Domingo, and C. Lopez-Galindez.** 1999. Drastic fitness loss in human immunodeficiency virus type 1 upon serial bottleneck events. *J. Virol.* **73**:2745–2751.